

TU Wien



Business Informatics Group
Institut für Softwaretechnik und Interaktive Systeme

Content Wiederverwendung in verteilten Content Management Systemen

Diplomarbeit zur Erlangung des akademischen Grades eines Diplomingenieurs

eingereicht bei o. Univ.-Prof. Mag. Dipl.-Ing. Dr. Gerti Kappel
mitbetreuender Assistent: Dr. Gerhard Kramler

Andreas Staribacher
2011 Höbersdorf 108

Wien, 9. Mai 2006

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benützt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Wien, 9. Mai 2006

Danksagung

Ich möchte mich bei Herrn Heidinger und Herrn Stanek von der Firma Siemens recht herzlich bedanken, dass sie mir diese Diplomarbeit ermöglicht haben.

Kurzfassung

Content Management Systeme sind in Unternehmen feste Bestandteile beim Contenterstellungsprozess. Sie unterstützen Autoren beim Erstellen und Bearbeiten von Content durch grafische Benutzeroberflächen, einfache Bedienung und unzähligen Hilfestellungen. Da oft gleicher Content auf verschiedenen Webseiten angezeigt und nur einmal erstellt werden soll, setzen immer mehr Unternehmen Content Wiederverwendung ein. Die meisten Content Management Systeme beinhalten – oft recht einfache – Funktionen, um Content wieder zu verwenden. Doch wie kann Content in Unternehmen, wo mehrere Abteilungen Content erfassen und verschiedene Content Management Systeme im Einsatz sind, wieder verwendet werden?

Diese Arbeit befasst sich mit allgemeinen Aspekten von Content Management Systemen und Content Wiederverwendung, wie Content Wiederverwendung in verteilten Content Management Systemen realisiert werden kann und wie Content Wiederverwendung von derzeit am Markt erhältlichen Content Management Systemen unterstützt wird.

Abstract

Content management systems are fixed components in organisations on creating content. They assist authors with graphical user interfaces, easy handling and immense helping functions, when they create and edit content. Because the same content should be displayed on multiple websites, but created only once, content reuse is introduced in organisations. Most content management systems have out-of-the-box – mostly basic – functions to reuse content. But how can content be reused in organisations, where various departments create content and different content management systems are in use?

This thesis deals with common aspects of content management systems and content reuse, how content reuse is implemented in distributed content management systems and how content reuse is supported by current content management systems.

Inhaltsverzeichnis

Abbildungsverzeichnis

Kapitel 1

Einleitung

Aufgrund der Komplexität von Internetauftritten, der Forderung nach aktuellen Webinhalten und einer Vielzahl an technisch nicht versierten Contenterstellern sind Content Management Systeme im heutigen Internet nicht mehr wegzudenken. Content Management Systeme erleichtern den Contenterstellungsprozess durch grafische Benutzeroberflächen, einfache Bedienung und unzähligen Funktionen und Hilfestellungen. Das Kapitel „Content Management Systeme allgemein“ befasst sich mit diesen Funktionen, erläutert den Begriff Content Management System und erklärt die Arten von verteilten Content Management Systemen.

Da oft gleicher Content auf verschiedenen Webseiten angezeigt und nur einmal erstellt werden soll, um dadurch Änderungen und Aktualisierung nur einmal durchführen zu müssen, bieten immer mehr Content Management Systeme Möglichkeiten zur Wiederverwendung von Content an. Im nächsten Kapitel „Content Wiederverwendung allgemein“ wird allgemein erklärt, was Content Wiederverwendung bedeutet und welche Methoden und Varianten es gibt, um Content wieder zu verwenden.

Da Content in einem Unternehmen von verschiedenen Abteilungen und Autoren erstellt und verwendet wird, und diese Abteilungen meistens nicht zusammenarbeiten, führt das oft zu den unterschiedlichsten Ausprägungen an Content. Um diesen Content wieder verwenden zu können, muss er in eine einheitliche Form gebracht wer-

den. Das Kapitel „Content-Vereinheitlichung für die Wiederverwendung“ befasst sich mit Aspekten, wie Content beschaffen sein muss, damit er wieder verwendet werden kann.

Da es zwar sehr viele Content Management Systeme gibt, die meist rudimentäre Funktionen für Content Wiederverwendung bereitstellen, müssen in einem verteilten CMS sehr viele Voraussetzungen erfüllt werden, damit Content wieder verwendet werden kann. Es müssen die unterschiedlichsten Anpassungen an den einzelnen Systemen durchgeführt werden, da kein einziges System die gewünschten Funktionen standardmäßig bereitstellt. Die Herausforderungen, die die Entwicklung eines solchen Systems beinhaltet, werden im Kapitel „Content Wiederverwendung in verteilten CMS“ dargestellt.

Das Kapitel „Bestehende CMS mit Funktionen für Content Wiederverwendung“ beinhaltet eine Evaluierung von aktuellen CM-Systemen. Da es eine unüberschaubare Menge an derzeit am Markt erhältlichen Content Management Systemen gibt, wurden ein paar der meist verwendeten Systeme für die Evaluierung herangezogen. Es wurden sowohl hochwertige Enterprise Content Management Systeme (Interwoven TeamSite, Microsoft Content Management Server) als auch Open-Source Produkte (OpenCms, Typo3) untersucht. Die Evaluierung beinhaltet allgemeine Informationen über die Content Management Systeme, eingesetzte Softwarekomponenten und Aspekte der Systemarchitektur. Im Speziellen wurden Möglichkeiten von Content Wiederverwendung aufgezeigt und diese genauer beschrieben.

Kapitel 2

Content Management Systeme allgemein

Die Website ist das Aushängeschild, die digitale Visitenkarte einer Firma im Internet. Deshalb ist es natürlich selbstverständlich, dass hier alles passen muss: Sei es nun der fehlerfreie Inhalt, funktionierende Links oder das einheitliche Corporate Design. Jedoch sind die Anforderungen an eine Website groß:

- Eine Website ist sehr komplex und besteht aus mehreren hundert einzelnen Seiten.
- Die Inhalte sollen häufig (z.B.: wöchentlich, täglich, stündlich) geändert werden.
- Mehrere Mitarbeiter sollen Inhalte bereitstellen.
- Diejenigen Mitarbeiter, die Inhalte bereitstellen sollen, haben keine oder geringe HTML- oder sonstige Programmierkenntnisse.
- Inhalte sollen im Rahmen eines mehrstufigen Entwicklungs- und Freigabe-Prozesses bereitgestellt werden.

Beim klassischen Web-Publishing ist ein Webmaster für die Einstellung der Inhalte in die Website verantwortlich. Er trägt in aufwendiger Koordinationsarbeit Informationen aus den Fachabteilungen zusammen, übernimmt die technische Aufbereitung und Programmierung und stellt die Inhalte online. Jedoch entspricht die Aufbereitung

nicht immer den Vorstellungen der Informationslieferanten. Auch Layoutänderungen ziehen einen erheblichen Nachteil nach sich: Eine große Menge an vorhandenen HTML-Seiten, Navigationsmenüs, etc. müssen manuell geändert oder erweitert werden. Inkonsistenzen und tote Links sind praktisch nicht zu vermeiden. Da dieser Aufwand für die Pflege und Wartung einer Website nicht gerade niedrig ist, müssen natürlich die Anforderungen an Aktualität, Qualität und Sicherheit kostengünstig erfüllt werden. Abhilfe schaffen Content Management Systeme (CMS). *[Ecom03]*

2.1 Was ist ein Content Management System?

Anhand seiner Wortbestandteile ist zunächst der Begriff „Content Management System" zu definieren:

2.1.1 Content

Als Content einer Website wird im Allgemeinen die Gesamtheit aller veröffentlichten digitalen Assets einer Website, wie Texte, Bilder, Grafiken und multimediale Dokumente, bezeichnet. Die einzelnen Assets werden jedoch nicht nur gesammelt, sondern müssen auch strukturiert und dargestellt werden. Text-Assets kann ein CMS zum Beispiel so strukturieren, dass genau spezifiziert wird, was eine Überschrift, was die Einleitung und was der Textkörper ist.
Ein innovatives Content Management System behandelt Content als Summe von wesentlichen Einzelinformationen. Diese sind: **Struktur, Darstellungsform und Inhalt**. Besonders wenn Informationen nicht nur für das menschliche Auge bestimmt sind, sondern zur automatisierten Weiterverarbeitung (Single Source Multiple Media) oder Weiterverwendung (Syndication) eingesetzt werden, führt dies zwangsläufig zu einer Zerlegung und separaten Datenhaltung dieser drei wesentlichen Bestandteile. In allen Dokumenten bzw. Informationsträgern ist diese Dreiteilung enthalten und ergibt sich nach folgendem Muster:

- Die **Struktur** ist eine inhaltliche Definition der Einzelinformationen und deren Abfolge bzw. Verschachtelung. Eine wissenschaftliche Arbeit besteht zum Beispiel aus Titel, Inhaltsverzeichnis, Literaturverzeichnis etc.

- Die **Darstellung** ist eine formale Beschreibung zur Repräsentation auf einem möglichen Ausgabemedium. Stylesheets enthalten informationstechnische Anweisungen, wie der Inhalt formatiert und positioniert werden soll. Es wird zum Beispiel ein Stylesheet verwendet um einen Inhalt auf einem A4 Blatt und ein weiteres um denselben Inhalt auf einer Webseite optisch darzustellen.
- Der **Inhalt** wird entsprechend der Strukturdefinitionen in Datenelementen abgebildet. Dabei ist die Datenhaltung des Inhalts von besonderer Bedeutung, denn die Metainformation muss erhalten bleiben.

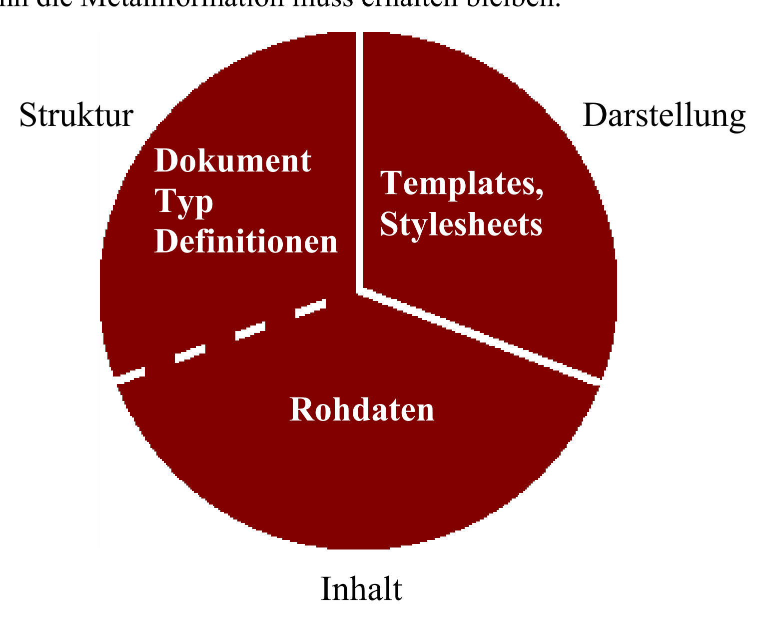


Abbildung 2.1 Content als Summe von Einzelinformationen
Quelle: [Schu00]

CMS, die Daten entsprechend dieser Dreiteilung behandeln, verfügen über vielfältige Möglichkeiten zur weiteren, clientseitigen und automatisierten Datenverarbeitung. Die Trennung von inhaltlichen und formalen Daten und der Einsatz von Stylesheets bietet neben der Darstellung von Informationen auf unterschiedlichen Ausgabemedien einen weiteren wichtigen Vorteil: Die Informationen können benutzer- und bedarfsgerecht aufbereitet werden. Durch geeignete Benutzerprofile können die personen- und prozessrelevanten Sichten auf Informationen realisiert werden. *[Schu00]*

2.1.2 Content Management

Der Begriff Content Management umfasst alle Aspekte der Verwaltung von Inhalten. Es stehen alle Facetten des „Managens“, also das Erstellen und Erfassen, das Speichern und Verfügbarmachen, das Präsentieren und Anzeigen sowie das Archivieren und dauerhafte Ablegen von Inhalten im Blickpunkt. *[Boik02]*

2.1.3 Content Management System

Ein Content Management System ist ein Datenbankgestütztes Software-System, das die abstrakte Content Management-Aufgabe in ganz bestimmter Weise mit programmtechnischen Mitteln zu lösen hilft. Meistens werden die mit einem CMS gemanagten Sites bzw. Inhalte als „dynamisch“ bezeichnet, während Websites, die auf normalen HTML Seiten basieren als „statisch“ bezeichnet werden. Da auch die Webseiten einer CMS-Site temporär als statische Seiten bereitgestellt werden, um z.B. die Performance zu verbessern oder die Systemlast zu verringern, sind die Übergänge aber fließend.

Bei einem CMS wird der Content in einer Datenbank erfasst und definiert. Ein spezielles Programm, meist ein so genannter Applikations-Server fragt dann die Inhalte aus der Datenbank ab, stellt sie zu einer Webseite zusammen und übermittelt diese dann an den Webserver. Der Webserver macht sie dann für den Internet-Benutzer erreichbar. Für ein CMS werden daher normalerweise zusätzlich ein Datenbankserver und ein Applikationsserver benötigt. Durch diese Technik soll der Prozess der Contentbereitstellung optimal unterstützt werden. Diejenigen, die Inhalte entwickeln und veröffentlichen wollen - z.B. Online-Redakteure, Produkt-Manager, usw. - brauchen nicht über HTML- oder sonstige Programmierkenntnisse verfügen. Sie erhalten ein einfaches Interface, z.B. einen Bildschirm mit Eingabemöglichkeiten, mit dem vergleichsweise einfach Inhalte erfasst und gespeichert werden können. Wenn ein Team an den Inhalten arbeitet, können Kollegen diese Inhalte vor der endgültigen Veröffentlichung betrachten und gegebenenfalls ändern. Am Ende dieses Inhalt-Entwicklungsprozesses kann der Inhalt dann für das Internet freigegeben werden. *[Ecom03]*

Hier eine Übersicht über diese Prozesse:

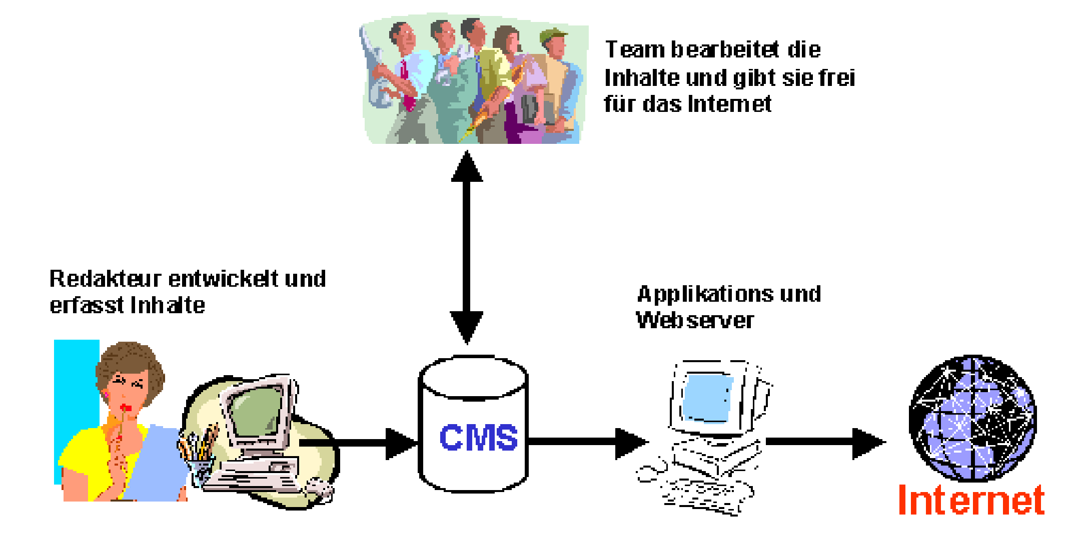


Abbildung 2.2 CMS-Prozesse
Quelle: [Ecom03]

2.2 Grundprinzip

Das Grundprinzip von Content Management Systemen ist die Trennung von Layout und Content. Der Aufbau einer Seite wird durch Templates gesteuert. Diese sind vordefinierte Grundgerüste für die Darstellung von Inhalten. Wird ein Dokument von einem Server abgerufen, werden die entsprechenden Bestandteile in ein Template geladen, dadurch angeordnet und dem Nutzer dargestellt. Für die Verwendung von Templates spricht vor allem die Möglichkeit, Informationen in gleich bleibendem Layout in großen Mengen darzustellen. Dabei ist es nicht jedes Mal nötig ein neues Dokument zu erstellen, sondern der Inhalt wird aus einer Datenbank ausgelesen, mittels Template formatiert und ausgegeben.

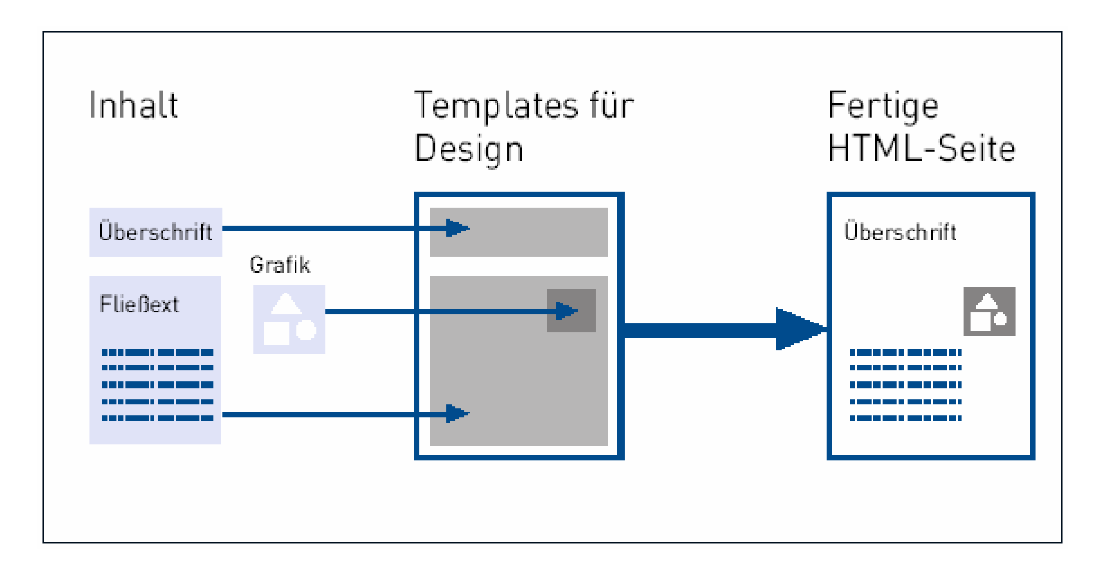


Abbildung 2.3 Trennung Inhalt-Layout
Quelle: [Pinu03]

Autoren können entweder mit ihren gewohnten Anwendungen arbeiten oder mit den integrierten Editoren, die leicht zu bedienen sind. Sie passen ihre Beiträge in das Template ein und sehen das Ergebnis sofort im Browser. Dadurch, dass Texte, Bilder und andere Bestandteile einer Website in einer Datenbank abgelegt und dort verwaltet werden, können diese auch mehrfach verwendet und verschiedene Versionen der Inhalte gespeichert werden. Ein späteres Zurückkehren zu einer bestimmten Version der Website ist somit zu jedem gewünschten Zeitpunkt gewährleistet. *[Pinu03]*

2.3 Kernkomponenten und Funktionen

Im Folgenden werden die fünf Kernkomponenten und erweiterte Funktionen, die von einem CMS unterstützt werden müssen, beschrieben. Zu den Kernkomponenten zählen das Assetmanagement, Workflowmanagement, Benutzer- und Zugriffsverwaltung, Import- und Exportschnittstelle und die Schnittstelle für funktionale Erweiterungen. Weitere Funktionen sind für ein CMS ebenso von Bedeutung, jedoch werden sie nicht ausnahmslos von allen Produktherstellern unterstützt. *[Knap03]*

2.3.1 Assetmanagement

Die zentrale Komponente jedes Content Management Systems ist das Assetmanagement, welches für die Verwaltung aller digitalen Assets verantwortlich ist. Getrennt von der letztendlichen Darstellung auf der Website wird Content, Idealerweise medienneutral, erfasst und gespeichert. In welcher Form die Speicherung und Ablage des Contents innerhalb eines CMS auch erfolgt, man spricht dabei immer vom Content Repository. Hier werden alle Assets, die auf der Website benutzt werden, zentral gespeichert. Der klare Vorteil zentraler Repositories liegt im konsistenten Datenstamm der von allen Mitarbeitern benutzt wird. Die Datensicherung und Administration des Systems wird, gegenüber stark verteilten Datenbeständen, durch die zentrale Speicherung erheblich erleichtert. In der heutigen Welt der CMS beherrschen vor allem zwei Modelle die Speicherung des Contents. Zum einen findet man die Ablage des Contents in Dateien und Verzeichnissen auf der eigenen Festplatte und zum anderen wird der Content in einer meist relationalen Datenbank abgelegt:

- **Dateien und Verzeichnisse**

  Informationen wurden seit jeher in Dateien gebündelt und diese wiederum strukturiert in Verzeichnissen abgelegt. Auf Webservern wird dieses Prinzip weiterhin angewendet, so hat jede Seite, die über das Web abgerufen wird, seine eigene Adresse - die URL, die nichts anderes als den Pfad auf dem Webserver darstellt auf dem sie angelegt ist. Für Dateien verschiedenster Formate gibt es eine breite Palette von Editoren, um den Content zu bearbeiten; so kann der Content in einem Dateisystem mit bestehenden Tools bearbeitet werden, was den Mitarbeitern den Einstieg in ein CMS erleichtert, da sie auf ihnen bekannte Werkzeuge zurückgreifen können. Dateien können auch ausführbaren Code beinhalten, d.h. sie sind teilweise oder komplett ein Programm oder Skript. Die Integration und Verwaltung von in Perl, ASP oder PHP geschriebenen Skripten und Programmen ist über die Dateistruktur sehr einfach. Neben den Assets werden auch die Templates im Dateisystem angelegt. Der Content lässt sich wie oben erwähnt mit einem beliebigen Editor bearbeiten, womit fast alle Freiheitsgrade der Gestaltung in HTML zu Verfügung stehen - mit dem Resultat flexibler Designs.

- **Datenbanken**
  Zum Zweck der Strukturierung von Informationen ist der Einsatz von Datenbanken bei der Speicherung des Contents unerlässlich. Für die ersten Schritte bei der Einführung eines CMS mittels Einsatzes einer Datenbank ist es nötig die Datenstrukturen zu definieren, d.h., die darzustellenden Informationen müssen in Teile zerlegt und für jede Informationsart eine Struktur definiert werden. So haben z.B. Pressemeldungen einen anderen Aufbau auf der Website als Stellenbeschreibungen oder Mitarbeiterportraits. Für jede Contentart ist eine leere Hülle in der Datenbank zu erstellen. Dies ist ein sehr wichtiger und entscheidender Schritt, da das Datenbankdesign lange vorhalten sollte und die Strukturen für einen langen Zeitraum festgelegt werden. Durch die Strukturierung des Contents ist dieser teilweise oder ganz auf der Seite verwendbar (Wiederverwendung), jedoch werden spezielle Tools benötigt, um auf den Content in der Datenbank zugreifen zu können.

Allgemein stellt sich die Frage, welches Konzept insgesamt gesehen wohl das bessere ist. Da beide Konzepte ihre Vor- und Nachteile besitzen, bedienen sich viele Hersteller von CMS beiden. Das Dateisystem wird weiterhin für wenig strukturierte und unstrukturierte Informationen genutzt, wobei dies nicht nur gestaltungsaufwendige HTML-Seiten betrifft, sondern ebenso Grafiken sowie Audio- und Video-Files. Auch Skripte, die auf dem Webserver ausgeführt werden, findet man weiterhin im Dateisystem. Die Datenbank hingegen wird zur Verwaltung der zu jedem Asset gehörenden Metainformationen verwendet sowie für Informationen und Textelemente, die man leicht strukturieren kann und mehrfach verschieden aufbereitet nutzen möchte. Dieser Ansatz hilft, eine flexible und offene Gestaltung beizubehalten und jeweils dort zu dynamisieren und zu strukturieren, wo es sinnvoll erscheint. *[Knap03]*

2.3.2 Workflowmanagement

Eine große Website lässt sich nur mit genau definierten Aufgaben jedes einzelnen Mitarbeiters effektiv verwalten. Dafür sind Mechanismen notwendig, die den Workflow (Ablauf der Arbeitsschritte) der Website im Content Management System abbilden. Ein Rollenkonzept regelt hierbei den aufgabenbezogenen Zugriff auf Teile des Systems. Das Rollenkonzept bildet die Arbeitsteilung eines Teams innerhalb des

Systems ab. Damit kann die Rechteverteilung intuitiv in nachvollziehbaren Rollen geschehen. Für jede Rolle wird ein Berechtigungskonzept erstellt, dass der entsprechenden Gruppe von Mitarbeitern übertragen wird. Meldet sich ein Mitarbeiter am Content Management System mit seiner Kennung und Passwort an, bekommt er Zugriff auf das System nur in dem Umfang, wie ihm dafür Berechtigungen entsprechend seiner Rolle übertragen wurden. Ein Administrator bekommt so z.B. alle Rechte für den Zugriff auf Systemeinstellungen, der Chefredakteur Zugriff auf die Sitestruktur und der Sportredakteur nur auf die Sportseiten. Mit einem Freigabezyklus für Publikationen wird festgelegt welche Instanzen diese bis zur Veröffentlichung durchlaufen muss. Wie in einem normalen Redaktionsprozess wird ein vom Redakteur erstellter Artikel der Chefredaktion vorgelegt, diese akzeptiert den Artikel oder gibt ihn zur Nachbearbeitung und Wiedervorlage an den Redakteur zurück. Content Management Systeme setzen diesen Vorgang in eine Automation um, die sich nicht nur auf einzelne Publikationen sondern auch auf Komponenten wie Grafiken etc. bezieht. Die entsprechenden Nachrichten werden selbständig durch das Content Management System generiert (E-Mail oder systeminterne Nachrichten). Da in einer Teamorientierten Umgebung gewährleistet sein muss, dass durch die Vielzahl von Mitarbeitern an einer Website keine Konflikte beim gemeinsamen Zugriff auf Publikationen oder Komponenten auftreten, wählt ein Redakteur eine zu bearbeitende Teilpublikation aus und das CMS sperrt z.B. für die Zeit der Bearbeitung den Zugriff für andere Nutzer auf diesen Teil der Publikation. Somit wird sichergestellt, dass keine Nutzerkonflikte bei zeitgleichen Zugriffen auf Dokumente stattfinden. *[Knap03]*

2.3.3 Benutzer- und Zugriffsverwaltung

Benutzer- und Zugriffsverwaltung arbeiten sehr eng zusammen. Die über die Benutzerverwaltung erfassten oder aus anderen Systemen importierten Benutzer erhalten entsprechend ihrer Rolle im Workflow Rechte im System. Diese können speziell für den Grafiker oder den Sportredakteur ausgelegt sein. Da Websites mit dem Vordringen der Internettechnologie in die verschiedensten Bereiche immer unternehmenskritischere Anwendungen werden, ist guter Zugriffsschutz in einem Content Management System eine Grundvoraussetzung. Somit wird ein Sicherheitskonzept benötigt, dass nach Außen absichert, intern aber die nötigen Zugriffe erlaubt. Der Schutz nach

Außen erfolgt über die Zugriffsverwaltung. Das Gros der CMS ist so angelegt, dass die Clients mit dem Server über das HTTP- bzw. HTTPS-Protokoll, welches mehr Sicherheit bietet, kommunizieren. Der Nutzer gelangt zu einer Loginseite bzw. zum Client des CMS und muss sich hier mittels Eingabe seiner Kennung und Passwort authentifizieren. Über die Authentifizierung des Nutzers lassen sich seine Berechtigungen (Benutzerrechte) im System ableiten. Für CMS sind zwei Ansätze der auf den Benutzer abgestimmten Darstellung auf Funktionen und Inhalte im System möglich:

- Darstellung aller Möglichkeiten (Rechte werden bei Aktionen geprüft)
- Darstellung vordefinierter Funktionen und Contents

Die erste Variante ist eher eine rudimentäre Umsetzung des Rechtekonzeptes in ein Bedienungskonzept. Der Nutzer bekommt alle Funktionen und Inhalte dargestellt und erst das Auslösen einer Aktion lässt seine Berechtigung überprüfen. In der zweiten Variante orientiert sich die Bedienung des Systems an den Rechten des Nutzers. Auf diese Weise bekommt der Anwender einen für ihn übersichtlicheren Client zur Hand, was die Benutzung des Systems deutlich erleichtert. *[Knap03]*

2.3.4 Import- und Exportschnittstelle

Das CMS ist die Basis für den Content der für eine Website verwaltet werden muss. Neben Informationen die darin speziell für das Web aufbereitet werden, existieren in Unternehmen weitere Informationen die aus unterschiedlichen Applikationen stammen. Ein CMS muss diese Informationen zur webgerechten Darstellung, wie auch für Exports, nutzen können, wobei folgendes Beachtung finden sollte: *[Knap03]*

- Import und Umwandlung unterschiedlicher digitaler Dokumentformate
- Einbindung von Informationen aus Datenbanken
- Import bestehender Websites
- Verarbeitung fremder Medientypen
- Offen für unterschiedliche Dokumentformate

2.3.5 Schnittstelle für funktionale Erweiterungen

Das Application Programming Interface (API) bezeichnet eine dokumentierte Programmierschnittstelle, die zur Erweiterung oder Anpassung durch Entwickler benutzt werden kann. *[Knap03]*

2.3.6 Erweiterte Funktionen

- **Mehrsprachigkeit**
  Bei vielen internationalen Unternehmen ist es erforderlich, Inhalte in mindestens zwei Sprachen zu publizieren.
- **Personalisierung**
  Explizite (manuell konfigurierbar) und die darüber hinausgehende implizite Personalisierung (zieht automatisch Rückschlüsse auf die Interessen des Informationsnachfragers) sind sehr wichtig, um die Website, speziell natürlich den Content auf die Wünsche der Nachfrager abzustimmen.
- **Caching von Inhalten**
  Hierbei lassen sich statische Seiten im CMS bereithalten, sodass nicht bei jedem Zugriff eine Datenbankabfrage oder eine Datentransformation erfolgen muss.
- **Generieren statischer Informationen**
  Periodische Aktualisierung aus Datenquellen und das Generieren von statischen Inhalten sind sehr wichtig für die Performance eines CMS.
- **Automatismen**
  Automatismen erstellen Sitemaps, kontextbezogene Navigationselemente oder Indexseiten zu den neuesten Inhalten.
- **Cross-Media-Publishing**
  Es sollen nicht nur Internet-Browser als „Kunden“ in Betracht gezogen werden, sondern es soll auch eine Vielzahl weiterer Schnittstellen (WAP, etc.) und Formate realisierbar sein.
- **Versionierung**
  Durch die zentrale Ablage der Daten ist eine effiziente Versionsverwaltung, um darzustellen, wann, welche Version im Web war, möglich. Ein späteres

Zurückkehren zu einer bestimmten Version der Website ist somit gewährleistet.

- **Link-Checking**
  Ein CMS überprüft bei jedem Bearbeitungsschritt die Verknüpfungen der Assets untereinander und somit die Konsistenz der Links. Stellt das System eine Inkonsistenz innerhalb der Verlinkung fest wird der Anwender darüber informiert. Werden Objekte in der Site verschoben, sollte ein CMS fähig sein, die zu ändernden Links nachzuvollziehen, umzustellen und bei einem Wegfall von Objekten auf die dabei entstehenden toten Links hinzuweisen bzw. diese aus den einzelnen Seiten zu entfernen. Veränderungen externer Links stellen insofern ein Problem dar, dass hier zwar eine Benachrichtigung an das System erfolgen, jedoch eine Aktualisierung nicht automatisch realisiert werden kann. *[Boik02]*

2.4 Verwandte Schlagworte und Abgrenzung

Web Content Management System (WCMS)

Web-Content-Management-Systeme, wie sie heute vielfach zum Aufbereiten von Inhalten im Web zum Einsatz kommen, zielen nur auf einen sehr kleinen Teil, eben die Präsentation im Internet, des dargestellten Kernprozesses ab. Gehen die Anforderungen weiter, indem auch andere Präsentationskanäle unterstützt werden sollen oder wenn die Wiederverwendung von Content eine Rolle spielt, greifen Web-CM-Systeme im Allgemeinen zu kurz. Darüber hinaus sind die integrierten Redaktionskomponenten in der Regel in ihrer Funktionalität zu beschränkt, um auch komplexe Produkte zu generieren, die Inhalte aller Medientypen enthalten. *[Baum01]*

Enterprise Content Management System (ECMS)

Hierbei handelt es sich um anspruchsvolle High-End-Lösungen mit starker Integration in bestehende Back-End-Systeme (z.B. IT Infrastruktur der Kunden und ERP Systeme). *[Berg03]*

Redaktionssystem

Der Begriff des Redaktionssystems wird immer wieder in Zusammenhang mit CMS benutzt oder als Synonym bzw. deutsche Entsprechung des Begriffes CMS gesehen. Redaktionssysteme sind jedoch eine Klasse von CMS, die eher für das Veröffentlichen von Inhalten ausgerichtet sind und nur in begrenztem Maße alle Komponenten eines CMS abdecken. *[Knap03]*

Document Management System (DMS)

In Unternehmen entstehen verschiedenste Dokumente, die Informationen erfassen oder bestimmte Vorgänge dokumentieren. Systeme für das Dokumentenmanagement helfen bei der Verwaltung und dem kontrollierten Zugriff auf Content, wobei die wesentlichen Features die Erfassung, Versionsvergabe oder Archivierung derselben sind. Verwendete Texte, Bilder und andere Bestandteile einer Website werden in einer Datenbank abgelegt, dort verwaltet und können mehrfach verwendet und in verschiedenen Versionen gespeichert werden. *[Knap03]*

Knowledge Management System (KMS)

Knowlegdemanagement versucht aus einmal erfassten Informationen Wissen zu generieren. Die Vielgestaltigkeit der Knowledgemanagement-Lösungen zeigt, dass es beim Versuch, Wissen aus strukturierten Informationen zu gewinnen und zusammenzuführen, die verschiedensten Ansätze gibt. Dabei geht es weniger um den Einsatz einer Technik, als vielmehr darum das Bewusstsein von Einzelnen oder Gruppen für diesen Prozess zu schärfen. Ein Intranet mag wenig nutzen, wenn der Anteil der Nutzer im Verhältnis zur Mitarbeiterzahl gering ist. Software für Knowledgemanagement-Systeme spezialisieren sich somit auf Teilbereiche, wie das intelligente Auffinden von Informationen oder den gruppenbasierten Wissensaustausch. Ein CMS kann z.B. eine Lösung, Knowledgemanagement in einem Unternehmen einzuführen, unterstützen. *[Knap03]*

Media Asset Management System (MAMS)

Die junge Disziplin Media Asset Management, oder das synonym verwendete Digital Asset Management, hat sich aus der Situation heraus entwickelt, dass es neben Web Content in den Unternehmen noch eine Vielzahl von anderen Inhalten mit spezifi-

schen Anforderungen gibt, die intelligent strukturiert und verwaltet, rasch aufbereitet und aktualisiert werden wollen (Bilder, Grafiken, Audio- und Video-Medienbausteine etc.). Innovative Konzepte und die Grundcharakteristiken des Internet (Autonomie, Offenheit, Transparenz, gleichberechtigte Beeinflussung durch Sender und Empfänger,…) haben dazu geführt, dass in diesem Umfeld Informationsasymmetrien weitgehend eliminiert werden können. Die Abhängigkeit von Spezialisten kann reduziert, der eigene Handlungsspielraum erhöht werden. Zudem hat das gedruckte Printmedium, allen Unkenrufen zum trotz, kaum an Stellenwert eingebüsst. Gerade aus der graphischen Industrie und aus der Medienindustrie haben sich viel versprechende Konzepte entwickelt, die gewaltigen Mengen an verschiedensten Inhalten adäquat zu verwalten und wieder zu verwenden. Als Vorläufer zu Media Asset Management Systemen können beispielsweise Bildkatalog-Systeme bezeichnet werden. Media Asset Management positioniert sich als System zur medienneutralen Strukturierung und Verwaltung von Medienobjekten mit starkem Marketingkontext. Das Konzept der Medienneutralität erlaubt die effiziente Bedienung verschiedenster Absatz- und Medienkanäle mit Inhalten und ist demnach, im Vergleich zu Web Content Management Systemen nicht auf spezifische Medientypen oder Medienkanäle fokussiert. Media Asset Management Systeme streben grundsätzlich nicht danach, Web Content Management Systeme (oder andere) zu verdrängen, vielmehr müssen MAM-Systeme äußerst offen angelegt sein, um mit anderen Systemen Daten austauschen zu können. *[Klei02]*

2.4.1 Abgrenzung CMS - DMS - KMS

Am Markt herrscht eine beträchtliche Unklarheit über den Unterschied zwischen Content Management Systemen, Dokument Management Systemen und Knowledge Management Systemen. Das kommt natürlich auch daher, dass viele Hersteller ihr Produkt so breit gefächert wie möglich am Markt positionieren wollen, und somit von allen drei Systemen Funktionen in ihren Produkten verwenden.

Primär geht es in einem DMS um die Unterstützung und Kontrolle des Lebenszyklus von Dokumenten. Wobei das Speichern und Archivieren einer großen Anzahl von Dokumenten im Vordergrund steht. Jedoch auch Funktionalitäten wie Versionierung, Check-in/Check-out von Dateien und Zugriffskontrolle sind Bestandteile von solchen Systemen. Bei einem KMS liegt der Fokus allerdings darauf, vorhandene Un-

ternehmensinformationen (z.B. Umsatz- und Absatzkennzahlen, aber insbesondere Erfahrungen) zu sammeln und zu verdichten, um Mitarbeiter bei Entscheidungsprozessen zu unterstützen. Da Wissen sehr wichtig für Unternehmen ist („Intellektuelles Kapital“) und auch vergänglich ist (Mitarbeiterfluktuation, Halbwertszeit) sind Firmen bemüht, ihre Informationen in einem KMS so strukturiert festzuhalten, dass sie auch leicht wieder gefunden werden können.

Somit liegt der Unterschied zwischen einem DMS und einem KMS auf der Hand: Ein DMS verwaltet ganze Dokumente (meist Dateien) zwar strukturiert, jedoch die einzelnen Informationen in den Dokumenten werden außer Acht gelassen. Hier setzt ein KMS auf, denn jegliche Form von Wissen und Informationen wird hier gegliedert festgehalten. Folglich erfolgt das Verwalten bei einem KMS mindestens eine Ebene tiefer als bei einem DMS. Zwischen einem CMS und einem DMS gibt es Überlappungen zum Beispiel bei Versionierung, Zugriffskontrolle und Workflow. Jedoch ein großer Unterschied besteht in der Verbindung der Dokumente zueinander. Wobei Dokumente in einem DMS untereinander fast nicht in Zusammenhang stehen, sind Informationen hingegen, die von einem CMS verwaltet werden, sehr stark miteinander verbunden. Ein KMS hat mit einem CMS im Grunde nur eines gemeinsam: strukturierte Datenhaltung. Ein KMS muss strukturiert aufgebaut sein, um Informationen schnell wieder zu finden, genauso wichtig ist in einem CMS das strukturierte Speichern der Assets. Gerade im Hinblick auf Wiederverwendung ist eine genaue Datenstruktur unabkömmlich. Abschließend ist zu sagen, dass sich diese drei Systeme nicht konkurrieren, sondern sie sollten sich gegenseitig ergänzen.

2.5 Verteilte Content Management Systeme

2.5.1 Zentrales CMS mit dezentraler Contenterfassung

Bei einem zentralen CMS mit dezentraler Contenterfassung gibt es genau ein zentrales Content Repository, in dem alle Inhalte gespeichert werden. Die Contenterfassung erfolgt jedoch dezentral. Das heißt jeglicher Inhalt wird von einer beliebigen Anzahl von Organisationseinheiten erfasst. Somit ist jede Einheit für ihren Content verantwortlich und einzig und allein zu Änderungen berechtigt.

Im Gegensatz dazu, bei einem komplett zentralen CMS, gibt es einige wenige Mitarbeiter (Administratoren, Webmaster, Sitemanager, etc.), die sich mit der Aktualisierung und redaktionellen Pflege des Contents und der Seiten befassen. Jedoch besteht eine sehr enge Liaison zwischen diesem Content-Team und den Organisationseinheiten, die für die Inhalte verantwortlich sind. Denn alle Informationen, die von dem Team veröffentlicht werden, müssen von der jeweiligen Organisationseinheit überprüft und freigegeben werden, um Fehlerfreiheit und Relevanz gewährleisten zu können. *[Robe02]*

Die Vorteile der dezentralen Contenterfassung liegen auf der Hand:

- Das Bemühen von vielen Autoren wirkt sich positiv auf die Inhalte aus.
- Die Kosten für die Contenterfassung und für Ressourcen sind auf das gesamte Unternehmen verteilt.
- Jeder Autor ist für seine Inhalte selbst verantwortlich.
- Die Contenterstellung ist in die täglichen Aktivitäten des Unternehmens integriert.
- Inhalte können schneller online gehen und sind somit immer up-to-date.

Natürlich zieht dieser Ansatz auch Nachteile mit sich:

- Einschulung von vielen Benutzern.
- Schwierigkeiten in der Motivation der Mitarbeiter, Inhalte einzupflegen.
- Die größte Schwierigkeit besteht jedoch darin, Zeit zu finden um Inhalte einzupflegen. Dadurch dass Mitarbeiter ganztägig mit dem Arbeitspensum ihrer normalen Aktivitäten ausgelastet sind, findet sich kaum Zeit zu zusätzlichen Tätigkeiten wie Inhalte einpflegen.

2.5.2 Verteiltes CMS

Ein verteiltes CMS unterscheidet sich im Grunde nur in einem Punkt von einem zentralen CMS mit dezentraler Contenterfassung: Das Content Repository wird nicht zentral verwaltet, sondern es gibt mehrere Content Repositories, in denen Inhalte gespeichert werden. Das heißt, jede Organisationseinheit ist mit einem eigenen CMS ausgestattet, und die dort eingepflegten Inhalte werden lokal im CMS gespeichert. Jedoch kann es eine von der „Dach-Organisationseinheit“ betriebene zentrale Website geben, in der alle Querschnittsinformationen und -funktionen über die einzelnen dezentralen, (teil-) autonomen Organisationseinheiten zusammenlaufen. Diese zentralen Websites werden in der Regel als „Portale“ bezeichnet.
In dezentralisierten Organisationsumgebungen muss ein CMS daher zusätzlich zur Erzeugung und Pflege von individuellen Einzel-Websites auch spezielle Integrationsleistungen für konsistente Portale erbringen. Die Spezialität dieser Leistungen liegt darin begründet, dass die Contentintegration verschiedener Einzel-Websites im laufenden Betrieb der Portale möglichst effizient, also automatisiert und mit nur begrenzten manuellen Eingriffen erfolgen soll. Es geht sogar so weit, dass ein CMS eine variable Anzahl von hierarchisch oder netzwerkartig geordneten Portalen zu einer geschlossenen Portal-Struktur zusammenführen kann. *[Schw04]*

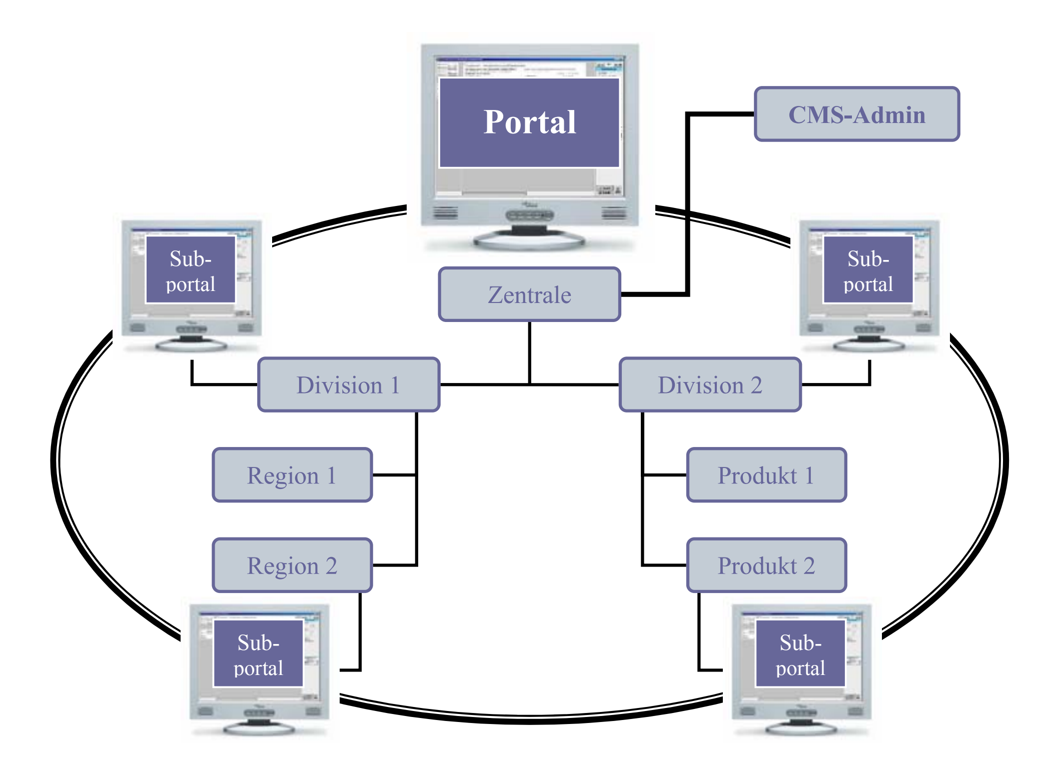


Abbildung 2.4 Dezentrale Unternehmensstruktur mit dezentralisierten Portalen
Quelle: [Schw04]

Nachteile gegenüber einem zentralen CMS mit dezentraler Contenterfassung ergeben sich eigentlich nur bei der Software: Erstens müssen sehr viele Arbeitsplatzrechner mit der CMS-Software konfiguriert werden, und zweitens steigen die Lizenzkosten der Software pro Arbeitsplatz.

Man unterscheidet 2 Arten von verteilten CMS:

- Verwendung des gleichen CMS-Produktes
  Bei allen Einzel-Websites wird dasselbe CMS-Produkt verwendet. Somit ist Content Syndication (= der Austausch von Inhalten zwischen Websites) sehr einfach, da alle CMS über dieselbe Export- und Importschnittstelle verfügen.
- Verwendung verschiedener CMS-Produkte
  Bei der Verwendung unterschiedlicher CMS-Produkte bei den Einzel-Websites ist Content Syndication schon weit schwieriger, da die vorhandenen Export- und Importschnittstellen meistens zueinander nicht kompatibel sind. So muss man oft eigene Integrationslösungen entwickeln, die auf offenen Standards, wie XML basieren sollten.

Kapitel 3

Content Wiederverwendung allgemein

Das gesamte Kapitel basiert auf [Rock03].

3.1 Was ist Content Wiederverwendung?

Content Wiederverwendung ist das Verfahren, vorhandene Contentkomponenten zu verwenden, um neue „Dokumente“ zu erstellen. Obwohl die Mehrheit von wieder verwendbarem Content Textorientiert ist, kann jeder mögliche Content wieder verwendet werden (wie Grafiken, Diagramme, Medien). Textbasierter Content ist am einfachsten wieder zu verwenden. So können Abschnitte, Absätze, Sätze oder sogar Wörter wieder verwendet werden. Es ist einfacher Grafiken, Diagramme und Medien in ihrer Gesamtheit wieder zu verwenden als Teile von ihnen. Jedoch neue Standards wie SVG (**S**calable **V**ector **G**raphics), ein neuer Internet Grafikstandard, erlauben es, auch wieder verwendbare Medien zu erstellen. Der XML Aspekt von SVG macht es möglich, alle Komponenten einer Webgrafik in Schichten zu unterteilen, und somit Wiederverwendung zu erleichtern.

Die meisten Unternehmen verwenden bereits Content wieder, obwohl sie ihn nur kopieren und einfügen (copy & paste). Das funktioniert so lange gut, solange der Content, und überall wo er wieder verwendet wird, nicht aktualisiert werden muss. Dann kann es sehr zeitintensiv werden, um jeden Ort zu finden, wo dieser Content verwendet und wieder verwendet wird, und ihn zu ändern. Nicht nur Zeit ist das

Problem, sondern auch Inkonsistenzen und Ungenauigkeiten sind vorprogrammiert. Das kann sogar soweit führen, dass am Ende zwei oder mehrere komplett unterschiedliche Contentquellen vorhanden sind.
Wiederverwendung kann gesehen werden als der Prozess der „Verlinkung“ zu einem Element, das wieder verwendbaren Content darstellt. Der wieder verwendbare Content wird zwar im Dokument angezeigt, jedoch gehört er nicht zum Dokument. Derzeit betreiben schon manche Unternehmen diese Art von Wiederverwendung mit Grafiken in Microsoft Word (Einfügen, Grafik, Aus Datei, Einfügen und Verknüpfen). Sobald das wieder verwendete Element aktualisiert wird, werden die Änderungen überall, wo dieses Element vorkommt, nachgezogen. Das erspart eine Menge an Zeit und Arbeit (Eine Änderung per Hand → Viele Änderungen automatisch).

3.2 Warum wird Content wieder verwendet?

Heutzutage können Unternehmen der Notwendigkeit nach schneller und Kundenorientierter Erstellung von Content kaum standhalten. Dies wird erschwert durch Auflagen wie Ressourcenverminderung, Zeit und Budget. Content Wiederverwendung kann zu einer drastischen Verbesserung führen, wie Content in einem Unternehmen erstellt wird. Die Verbesserungen umfassen höhere Qualität, Konsistenz, langfristig verringerte Zeiten und Kosten für Entwicklung und Wartung. Außerdem unterstützt Wiederverwendung bei kurzfristiger Umgestaltung von Content, beim Auffinden von Content und bei der Feststellung von Contentwünschen. Content Wiederverwendung bietet folgende Vorteile:

- Erhöhte Konsistenz
  Wenn Content einmal geschrieben und mehrere Male wieder verwendet worden ist, ist er konsistent, wo auch immer er eingesetzt ist. Diese Konsistenz führt zu einem qualitativ höheren Content.
- Verringerte Entwicklungs- und Wartungskosten
  Die Entwicklungskosten sind verringert, weil die Menge an Content, die ein Autor erstellen muss, verringert ist. Anstatt den gesamten Content neu zu schreiben oder sich die Zeit zu nehmen wieder verwendbaren Content zu finden und zu kopieren, ist wieder verwendbarer Content entweder schnell verfügbar durch verbesserte Managementmöglichkeiten (wie Metadaten oder

CMS), oder ist automatisch für Autoren sichtbar gemacht (systematische Wiederverwendung, siehe 3.4.2).

- Rasche Umgestaltung
  Wieder verwendbarer Content ist modular (Kleine Komponenten, die in Kombination mit anderen verwendet werden können). Modularer, wieder verwendbarer Content ist leichter umzugestalten, um Änderungsbedürfnissen nachzukommen. So kann man leicht die Reihenfolge der Module ändern, neue Module hinzufügen, bestehende Module entfernen und die Module dazu zu verwenden, um neue Informationen zu generieren.
- Übersetzung
  Die Kosten für Übersetzungen können durch Wiederverwendung erheblich gesenkt werden. Wenn Content wieder verwendet wird, so wird er auch nur einmal übersetzt und automatisch, überall wo er vorkommt, eingefügt. Somit müssen nur mehr die anderen Contentteile übersetzt werden.

3.3 Parallelität zu anderen Industriezweigen

Das Konzept der Wiederverwendung ist nicht neu. Viele Industrien haben Wiederverwendung eingeführt, um ihre Kosten zu senken, die Produktivität zu erhöhen und ihre Prozesse zu standardisieren. Im Folgenden werden Wiederverwendungskonzepte verschiedener Industriezweige vorgestellt.

Fertigungstechnik

Fertigungsfirmen erstellen nicht immer neue Komponenten ihres Produktes, wenn sie ein neues Produkt schaffen. Autos zum Beispiel unterscheiden sich eher im Aussehen als im Aufbau. Somit werden sehr viele Teile in verschiedenen Fabrikaten verwendet oder sogar in Fabrikaten von anderen Herstellern (wie Kugellager, Reifen oder Zündkerzen).

Softwareindustrie

Früher entwickelte die Softwareindustrie Software hauptsächlich wie heute Content erstellt wird, indem jeder Programmierer seinen eigenen Code schrieb und manchmal bestehenden Code kopierte und einfügte. Als nun das Konzept der Softwarewieder-

verwendung akzeptiert wurde, begannen Unternehmen Code zu entwickeln, der modular und speziell für Wiederverwendung konzipiert war.

Technische Publikationen

Getrieben durch das Verlangen von ein und derselben Information mehrere Versionen in verschiedenen Sprachen und unterschiedlichen Medien zu erstellen, entwickelte dieser Industriezweig eigene Wiederverwendungsstrategien. Sie verwendeten das Prinzip der „single source"-Wiederverwendung, das heißt, dass es eine einzige Quelle für den Content gibt. Content wird einmal geschrieben, an einem Ort gespeichert und mehrmals wieder verwendet.

Medien

Die Medien verwenden schon lange Content Wiederverwendung. Reporter schreiben Content einmal und Elemente dieser Originale werden in unterschiedlichen Formaten und Sprachen veröffentlicht. So kann es sein, dass Content eines Artikels in einer Fülle von Formaten vorkommt (in einer Zeitung, auf einer Website, auf PDAs und Mobiltelefonen, auf elektronischen Anschlagtafeln, und in verschiedenen Sprachen).

3.4 Methoden von Content Wiederverwendung

Es gibt zwei Methoden von Content Wiederverwendung:

- Opportunistische Wiederverwendung (opportunistic reuse)
  Ein Autor trifft die bewusste Entscheidung, Content wieder zu verwenden.
- Systematische Wiederverwendung (systematic reuse)
  Geplante Wiederverwendung, wo Content automatisch eingefügt wird.

Nachstehend werden diese Methoden ausführlich beschrieben.

3.4.1 Opportunistische Wiederverwendung

Opportunistische Wiederverwendung tritt auf, wenn ein Autor die bewusste Entscheidung trifft, ein Element zu finden, es abzurufen und es wieder zu verwenden. Diese Methode verlangt, dass Autoren die Möglichkeiten von Wiederverwendung kennen und auch motiviert sind, nach wieder verwendbarem Content zu suchen und ihn abzurufen. Opportunistische Wiederverwendung wird am häufigsten eingesetzt

und ist von keiner bestimmten Technologie abhängig. Diese Form der Wiederverwendung kann auch ohne ein CMS angewendet werden, jedoch ist ein CMS empfehlenswert. Viele Unternehmen verwenden opportunistische Wiederverwendung, wenn sie ihre Informationsprodukte schnell umgestalten wollen, um den Anforderungen an neue Produkte oder des Kunden zu entsprechen.
Jeder mögliche Content kann in so einer Situation verwendet werden. Irgendwie ist die opportunistische Wiederverwendung ein Ersatz für das „Kopieren & Einfügen", das in vielen Unternehmen angewendet wird. Dennoch ist opportunistische Wiederverwendung nicht Kopieren und Einfügen, da wieder verwendeter Content eigentlich nicht im Dokument selbst beinhaltet ist, sondern viel mehr ein Zeiger auf den ursprünglichen Content ist.
Opportunistische Wiederverwendung führt oft zu keiner Wiederverwendung, da die ganze Last den Autoren, die Content wieder verwenden wollen, auferlegt ist. Wenn die Motivation oder das Bewusstsein fehlt, dass ein passender wieder verwendbarer Content existieren könnte, oder es ist schwierig für Autoren, die relevanten Elemente zu finden, wird Wiederverwendung nicht angewendet.
Um die Effektivität von opportunistischer Wiederverwendung zu steigern, können Unternehmen das Auffinden von Informationen optimieren, Richtlinien bereitstellen und Autoren richtiggehend trainieren. Das Auffinden und Abrufen von Informationen kann optimiert werden durch den Einsatz von Content Management Systemen, effektiver Klassifizierung von Content und wirkungsvollen Metadaten. Richtlinien für Wiederverwendung beinhalten Modelle, um Autoren zu leiten, wo Content wieder verwendet werden sollte. Schulungen garantieren, dass Autoren Modelle und Content Management Systeme richtig verwenden und die Richtlinien befolgen.

Beispiel

Ein Autor erstellt eine Webseite für ein neues Produkt. Er weiß ganz genau, dass das Produkt sehr ähnlich zu einem bestehenden, schon dokumentierten Produkt ist. Er sucht im CMS nach Content für Standardeigenschaften. Er verwendet den passenden Content wieder.

3.4.2 Systematische Wiederverwendung

Systematische Wiederverwendung ist geplante Wiederverwendung. Spezieller Content wird als wieder verwendbar an bestimmten Positionen gekennzeichnet. Dann fügt das CMS an den entsprechenden Positionen im Dokument den wieder verwendbaren Content ein. Somit braucht sich der Autor nicht darum kümmern, ob der wieder verwendbare Content existiert und er muss nicht danach suchen oder ihn abrufen. Systematische Wiederverwendung stellt sicher, dass Content wieder verwendet wird und entlastet den Autor zu wissen, ob wieder verwendbarer Content existiert, wieder verwendbaren Content zu finden und ihn richtig einzufügen.

Systematische Wiederverwendung ist abhängig von detaillierten Informationsmodellen, Wiederverwendungsplänen und der Contentanlieferung durch das CMS. Das Content Management System benötigt das Informationsmodell, um Content, der wieder verwendet werden soll, zu identifizieren. Systematische Wiederverwendung ist sehr teuer zu implementieren, da sehr viel Zeit für das Erstellen von Informationsmodellen und Wiederverwendungsplänen und den Einsatz der angemessenen Technologie gebraucht wird, jedoch bietet es die größte Rendite (Return on Investment, ROI), da Wiederverwendung garantiert ist.

Systematische Wiederverwendung sollte angewendet werden, wenn Content sehr strukturiert ist und wenn man eindeutig feststellen kann, wann und wo Content wieder verwendet werden soll. Beispiele von systematischer Wiederverwendung können Standardfirmeninformationen wie Markenzeichen, Copyrights, Lizenzinformationen und Garantien beinhalten. Allerdings kann diese Form der Wiederverwendung immer dann verwendet werden, wo immer ein Unternehmen sicher gehen will, dass Content wieder verwendet werden soll (zum Beispiel, in Produktbeschreibungen, Ankündigungen, Regeln, Mitteilungen, Definitionen, Unternehmensprofilen, etc.).

Beispiel

Ein Consultingunternehmen, das Prozessumstrukturierungen durchführt, erstellt eine Reihe von Berichten (Analyse, Empfehlungen, Implementierung). Ein Analysebericht identifiziert die beobachteten Problemfälle im Unternehmen. Der Consultant fasst jeden Fall zu Beginn eines Abschnittes zusammen. Die Zusammenfassungen werden in der Kurzfassung unverändert wieder verwendet. Die Fälle werden in den Empfehlungsberichten wieder verwendet und gefolgt von einem Verbesserungsvor-

schlag niedergeschrieben. Die Empfehlungen werden zu Beginn jedes Abschnittes zusammengefasst und wiederum in der Kurzfassung wieder verwendet.

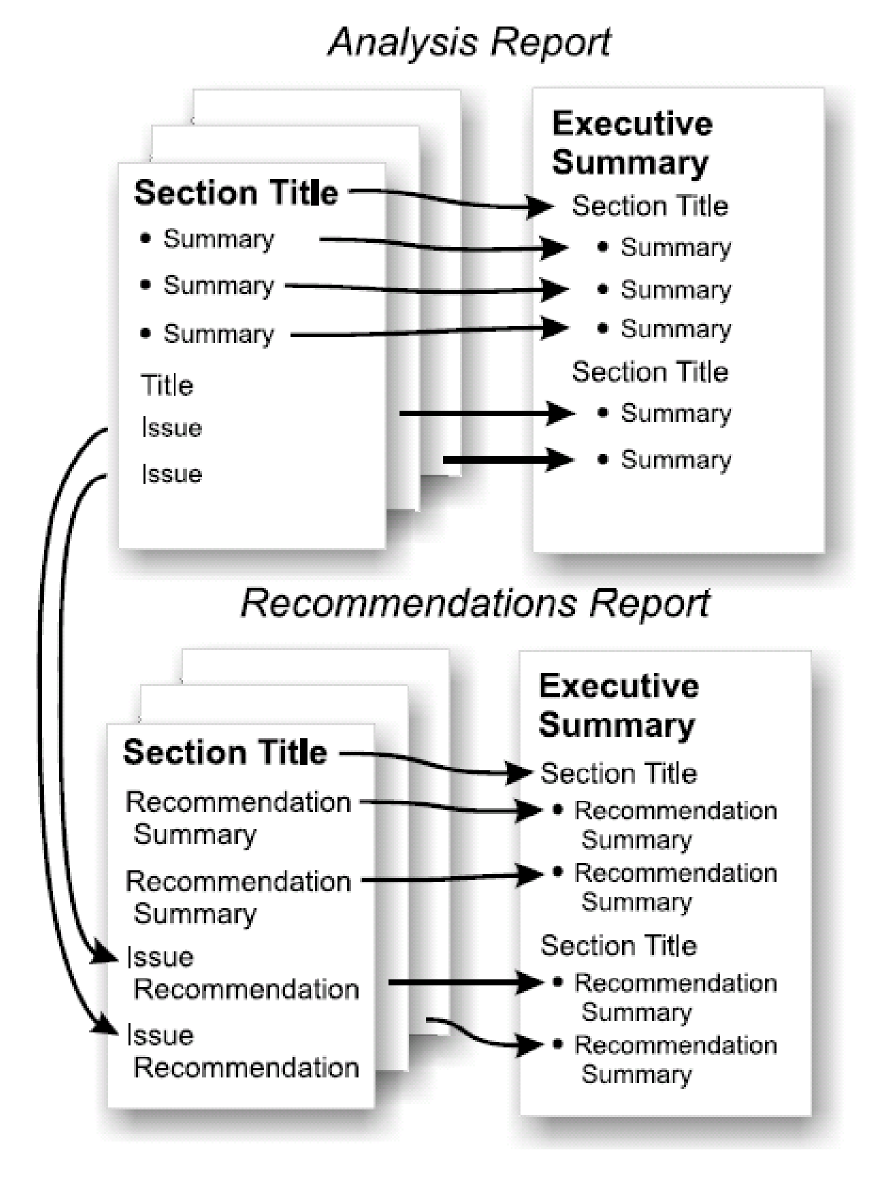


Abbildung 3.1 Systematische Wiederverwendung
Quelle: [Rock03]

3.5 Varianten von Content Wiederverwendung

Innerhalb jeder zuvor genannten Wiederverwendungsmethode gibt es drei verschiedene Varianten:

- Gesperrte Wiederverwendung (locked reuse)
  Wieder verwendeter Content kann nicht verändert werden
- Abgeleitete Wiederverwendung (derivative reuse)
  Wieder verwendeter Content kann verändert werden
- Verschachtelte Wiederverwendung (nested reuse)
  Unterschiedliche Versionen des Contents sind im selben Element enthalten

Im Folgenden werden die Varianten näher erläutert.

3.5.1 Gesperrte Wiederverwendung

Gesperrte Wiederverwendung tritt dann auf, wenn Content unverändert wieder verwendet wird. Der Content wird „gesperrt", damit Autoren ihn nicht verändern können. Nur Autoren mit den entsprechenden Rechten dürfen Änderungen an diesen Elementen vornehmen. Man kann gesperrte Elemente opportunistisch oder systematisch wieder verwenden.

Contenttypen, die im Allgemeinen gesperrt sind, umfassen Rechtsinformationen, Haftungsausschlüsse oder Markeninformationen. Jedoch kann jeder Content, der von anderen nicht geändert werden soll, gesperrt werden.

3.5.2 Abgeleitete Wiederverwendung

Wenn ein Autor ein wieder verwendbares Contentelement bearbeitet, ist das entstandene Element eine Ableitung, ein „Child"-Element des originalen „Parent"-Elements. Ein Element, das systematisch oder opportunistisch wieder verwendet wird und nicht „gesperrt" ist, wird zu einem abgeleiteten Element. Ein abgeleitetes Element ist weiterhin mit dem Originalelement verbunden, denn wenn das Originalelement verändert wird, wird der Autor des abgeleiteten Elementes benachrichtigt und die Änderungen können, jedoch müssen nicht eingearbeitet werden. Durch die Verwendung von abgeleiteter Wiederverwendung wird die Flexibilität der Wiederverwendung erhöht.

Abgeleitete Wiederverwendung wird in Unternehmen verwendet, wo Hauptcontent erhalten bleiben soll und folgende Änderungen am Content vorgenommen werden können:

- Zeitform
  Beispiel: Content wird in einem bestimmten Zeitraum in verschiedenen Berichten verwendet.
- Rechtschreibung
  Beispiel: amerikanisches im Gegensatz zu britischem Englisch
- Reihenfolge des Contents
- Betonung
  Beispiel: die unterschiedliche Gewichtung von Eigenschaften, um regionale Bedürfnisse abzudecken
- Verwendung
  Beispiel: Ein Werbeprospekt, das sehr Marketingorientiert ist im Gegensatz zu einem Benutzerhandbuch, das mehr Produktorientiert ist.

Übersetzter Content ist eine Sonderform der abgeleiteten Wiederverwendung (Das Element in der Originalsprache wird wieder verwendet, jedoch der Content wird durch den übersetzten Content ersetzt). Allerdings bleibt die Beziehung zwischen den Elementen erhalten, um sicherzustellen, dass der übersetzte Content geändert werden muss, wenn der Quellcontent verändert wurde. In diesem Fall wird der gesamte Content eines Elementes geändert, denn normalerweise werden bei der abgeleiteten Wiederverwendung nur einzelne Wörter geändert oder Sätze hinzugefügt oder gelöscht.

Beispiel

Ein Unternehmen plant die Durchführung eines Usability-Tests ihres neuen Internetauftritts. Sie erstellen einen Planungsbericht für das Management, um die Benutzertypen aufzuzeigen, die sie auswählen werden, dann führen sie den Test durch und schreiben eine Zusammenfassung. Sie verwenden Content des Planungsberichtes abgeleitet in der Zusammenfassung wieder. Abbildung 3.2 zeigt die Unterschiede der beiden Berichte auf:

<table>
<tr><th>Usability test proposal report
selecting test subjects</th><th>Usability summary report
selected test subjects</th></tr>
<tr><td>The following criteria for selection will be used to ensure we get appropriate test subjects from all our key customer areas including:

• Commercial customers
• General public customers
• Decision-makers

The selected test subjects should meet the following criteria:

• English as a first language
• Uses the Internet at least twice a week
• Uses Product ABC
• Has previously used our Help Desk for assistance

Each participant will be asked to complete a pre-test questionnaire to enable us to assess prior knowledge and experience.</td><td>The following criteria for selection were used to ensure we got appropriate test subjects from all our key customer areas including:

• Commercial customers
• General public customers
• Decision-makers

The selected test subjects met the following criteria:

• English as a first language
• Uses the Internet at least twice a week
• Uses Product ABC
• Has previously used our Help Desk for assistance

Each participant was asked to complete a pre-test questionnaire that enabled us to assess prior knowledge and experience.</td></tr>
</table>

Abbildung 3.2 Beispiel einer abgeleiteten Wiederverwendung
Quelle: [Rock03]

3.5.3 Verschachtelte Wiederverwendung

Verschachtelte Wiederverwendung tritt auf, wenn eine Anzahl von wieder verwendbaren Elementen in einem Element eingebettet (verschachtelt) ist. Die Summe aller Elemente erzeugt ein einziges Element, und jede Teilmenge kann in unterschiedlichen Informationsprodukten eingesetzt werden. Verschachtelte wieder verwendbare Informationen ermöglichen Autoren Content für alle Ausgabemöglichkeiten zur selben Zeit zu erstellen. Der gesamte relevante Content ist auf einen Blick verfügbar, wenn Content verschachtelt ist.

Verschachtelte Wiederverwendung kann eingesetzt werden, wann mehr oder weniger detaillierter Content für ein oder mehrere Informationsprodukte nötig ist, oder es ist für den Autor einfacher, jeden notwendigen Content zu einem Thema an einem Ort zu erstellen.

Beispiel

Ein Unternehmen stellt ein Produkt B-Brother her. Sie verwenden eine Produktbeschreibung in drei verschiedenen Informationsprodukten wieder: Werbeprospekt, Betriebshandbuch, Webshop.

Werbeprospekt	Betriebshandbuch	Webshop
The B-Brother model 1984 is a programmable device that connects directly to consumers' televisions to track the channels they flip to, what programs they record, and what commercials they skip. The information is instantly transmitted to the cable or satellite provider.	The B-Brother connects directly to consumers' televisions. It can be programmed to track what channels they flip to, what programs they record, and what commercials they skip. The information is transmitted to the cable or satellite provider.	The B-Brother model 1984 is a device that connects directly to consumers' televisions to track their television watching habits.

Abbildung 3.3 Vergleich von Produktbeschreibungscontent
Quelle: [Rock03]

Wie man in Abbildung 3.3 sehen kann, ist der Content der drei Beispiele manchmal zwar nur leicht unterschiedlich, jedoch nach wie vor unterschiedlich. Der Content muss nicht unterschiedlich sein. Zuerst müssen Autoren den Content vereinheitlichen (siehe Abbildung 3.4), um eine Produkterklärung zu erstellen, die das Produkt aus Sicht des Unternehmens genau beschreibt.

The B-Brother model 1984 is a device that connects directly to a consumer's television to track their television watching habits. It can be programmed to track the channels they flip to, what programs they record and what commercials they skip. The information is instantly transmitted to the cable or satellite provider.

Abbildung 3.4 Vereinheitlichte Produktbeschreibung
Quelle: [Rock03]

Dann wird genau identifiziert, welche Komponenten für jedes Informationsprodukt passend sind. (siehe Abbildung 3.5)

<The B-Brother model 1984 is a device that connects directly to a consumer's television to track their television watching habits.> *[Werbeprospekt, Betriebshandbuch, Webshop]*
<It can be programmed to track the channels they flip to, what programs they record and what commercials they skip. The information is instantly transmitted to the cable or satellite provider.> *[Werbeprospekt, Betriebshandbuch]*

Abbildung 3.5 Produktbeschreibung für Wiederverwendung markiert
Quelle: [Rock03]

Die Produktbeschreibung beinhaltet eine Kurzbeschreibung des Produktes (1. Satz), die in allen drei Informationsprodukten verwendet werden kann. Der 2. und 3. Satz können im Werbeprospekt und im Betriebshandbuch verwendet werden. (siehe Abbildung 3.6)

Werbeprospekt	Betriebshandbuch	Webshop
The B-Brother model 1984 is a device that connects directly to a consumer's television to track their television watching habits. It can be programmed to track the channels they flip to, what programs they record and what commercials they skip. The information is instantly transmitted to the cable or satellite provider.	The B-Brother model 1984 is a device that connects directly to a consumer's television to track their television watching habits. It can be programmed to track the channels they flip to, what programs they record and what commercials they skip. The information is instantly transmitted to the cable or satellite provider.	The B-Brother model 1984 is a device that connects directly to a consumer's television to track their television watching habits.

Abbildung 3.6 Wiederverwendung der Produktbeschreibung
Quelle: [Rock03]

3.6 Wann macht Wiederverwendung keinen Sinn?

Während man immer versuchen sollte so viel Content wie möglich wieder zu verwenden, gibt es Zeiten, wo Wiederverwendung nicht angebracht ist. Nicht jedes Stück Content ist wieder verwendbar, so sind Satzfragmente und einzelne Wörter ungeeignet für die Wiederverwendung. Je genauer die Contentelemente aufgeteilt werden, desto komplexer sind sie wieder zu verwenden. Um sicherzustellen, dass ein Element in jedem Fall wieder verwendet werden kann, sollten sehr generische Elemente verwendet werden. Generische wieder verwendbare Elemente erfüllen die Anforderungen an die Wiederverwendung, jedoch beeinträchtigen die Usability und Verständlichkeit des Contents.

Content sollte passend und effektiv wieder verwendet werden, jedoch darf die Qualität des Contents nicht beeinträchtigen werden und wieder verwendbarer Content schwer zu erstellen, zu finden und zu verwalten sein.

Kapitel 4

Content-Vereinheitlichung für die Wiederverwendung

Das gesamte Kapitel basiert auf [Rock03].

Content ist das Herzstück in einem verteilten CMS und wird in mehreren unabhängigen und oft sehr ungleichen Abteilungen erstellt und verwendet. Da die einzelnen Abteilungen bei der Contenterstellung meistens nicht miteinander zusammenarbeiten und die eine von der Arbeit der anderen nichts erfährt, führt dies oft zu Content der erstellt und wiedererstellt und wiedererstellt wird, oft mit den unterschiedlichsten Ausprägungen an Permutationen. Deswegen ist es unbedingt nötig, jeglichen Content in eine einheitliche Form zu bringen, indem man bei der Erstellung darauf achtet, dass ein konsistenter Schreibstil und eine geradlinige Struktur verwendet werden. Außerdem sollte Content so gespeichert werden, dass wer auch immer ihn benötigt, ihn findet, ihn abrufen kann, ihn wieder verwenden und an jeden ausliefern kann. Aus diesem Grund bedient man sich eines so genannten Informationsmodells, welches spezifiziert, wie Informationen wieder verwendet werden und wie sie geschrieben und strukturiert werden müssen, um Wiederverwendung (Reuse) überhaupt möglich machen zu können. Diese Modelle liefern das Framework, das Autoren, Rezensenten und Architekten bei der Erstellung, bei der Nachbearbeitung und beim Publizieren von Content leitet. Im Nachstehenden wird die Informationsmodellierung näher erläutert.

4.1 Erstellen von Informationsmodellen

4.1.1 Allgemein

Bevor man Content modellieren kann bzw. ein Informationsmodell erstellen kann, muss man zuerst herausfinden, wie Content in einer Organisation genutzt wird. Durch eine so genannte Content-Analyse entwickelt man ein Verständnis für verschiedenste Arten von Informationen, wer welche Informationen nutzt und in welchem Kontext sie verwendet werden. Dabei ist jedoch das Wichtigste, dass man sieht, wie Informationen wieder verwendet werden können, um die verschiedensten Bereiche abzudecken.

Hat man zum Beispiel durch eine Content-Analyse herausgefunden, dass man eine Produktbeschreibung an 20 verschiedenen Stellen publiziert hat, die von 10 verschiedenen Personen in 5 unterschiedlichen Abteilungen geschrieben wurden (vielleicht sogar noch in 7 verschiedenen Sprachen), so stellt man natürlich fest, dass viele Produktbeschreibungen sinngemäß nicht übereinstimmen, verschiedene Informationstypen beinhalten (manche beinhalten Kosteninformationen, andere nicht) und in unterschiedlichsten Weisen strukturiert sind, sodass Benutzer sich schwer tun, sie als Produktbeschreibungen zu identifizieren. Ferner bringen Autoren getrennt von anderen ihre Beschreibungen auf den neuesten Stand, sodass andere Nutzer derselben Beschreibung gleichzeitig über die Änderungen nicht informiert werden und somit Inkonsistenzen vorprogrammiert sind.

Während der Informationsmodellierung wird eine Strategie entwickelt, um Produktbeschreibungen zu vereinheitlichen, damit sie, wo auch immer sie auftauchen, konsistent sind, dieselben Informationstypen enthalten und gleich strukturiert sind. Modelle spiegeln die Struktur von Informationsprodukten wider, deswegen wird im folgenden Abschnitt genau erläutert, was Struktur ist.

4.1.2 Struktur

Informationsprodukte wie Produktbeschreibungen, Presseerklärungen, Briefe, Kataloge, usw. besitzen eine deutlich erkennbare Struktur, die wiederholt verwendet wird, immer wenn eines dieser Informationsprodukte erstellt wird. Zum Beispiel wie aus Abbildung 4.1 ersichtlich, hat ein Brief eine reproduzierbare Struktur.

Datum (Tag, Monat, Jahr)	*1. September 2005*
Name (Vor-, Nachname) **Adresse (Firmenname, Adresse, PLZ, Ort, Staat)**	*Vorname Nachname* *Firmenname* *Adresse* *PLZ Ort* *Staat*
Anrede	*Sehr geehrter Herr/Frau Nachname,*
Betreff	*AW: Ridebis, et licet rideas*
Absatz	*Ridebis, et licet rideas. Ego ille, quem nosti, apros tres et quidem pulcherrimos cepi. 'Ipse?' inquis. Ipse, non tamen ut omnino ab inertia mea et quiete discederem. Mirum est, ut animus agitatione motuque corporis excitetur.*
Aufzählung	• *Non est, quod contemnas hoc studendi genus.* • *Mirum est, ut animus agitatione motuque corporis excitetur.*
Überschrift	*Undique*
Absatz	*Iam undique silvae et solitudo ipsumque illud silentium, quod quod venationi datur, magna cogitationis incitamenta sunt.*
Nummerierung	*1. Proinde, cum vanabere, licebit auctore me ut panarium et laguncularn sic etiam pugillares feras.* *2. Ridebis, et licet rideas*
Absatz	*Ego ille, quem nosti, apros tres et quidem pulcherrimos cepi. 'Ipse?' inquis. Ipse, non tamen ut omnino ab inertia mea et quiete discederem.*
Abschluss (Autor, Berufstitel)	*Hochachtungsvoll* *Andreas Staribacher* *Student*

Abbildung 4.1 Struktur eines Briefes
Quelle: [Rock03]

Dieser Brief hat folgende Strukturelemente:

- Datum
- Adresse
- Anrede
- Betreff
- Hauptteil
- Abschluss

Diese Elemente können noch weiter unterteilt werden:

- Datum
  - Tag
  - Monat
  - Jahr
- Adresse
  - Name
    - Vorname
    - Nachname
  - Straße
  - Postleitzahl
  - Ort
  - Staat
- Anrede
  - Sehr geehrter Herr/Frau
  - Nachname
- Betreff
  - AW:
  - Thema
- Hauptteil
- Abschluss
  - Autor
  - Berufstitel

Das Unterteilen des Hauptteils ist jedoch sehr schwierig, da er wenig strukturiert ist. Autoren können die Struktur des Hauptteils auf verschiedene Arten abhängig von ihren Stilen, Content-Beschaffenheit und die Wirkung, die sie erzielen möchten wählen.

Der Hauptteil kann nur in sehr generische Elemente, die in beliebiger Reihenfolge verwendet werden können, unterteilt werden:

- Hauptteil
    - Absätze
    - Aufzählungen
    - Nummerierungen
    - Überschrift

Die Strukturierung des Contents ist ein sehr kritischer Prozess, da der gesamte Content einer Organisation ohne Rücksicht auf den, der ihn erstellt, vereinheitlicht wird. Das Maß, in dem Content strukturiert wird, hängt von der Granularität des Informationsmaterials ab und wird im Folgenden näher erklärt.

4.1.3 Granularität

Die Granularität – oder auch Detailgrad – bestimmt das kleinste Stück von Information, das wieder verwendbar ist. Granularität muss im Informationsmodell reflektiert werden, obwohl der Level der Granularität im gesamten Content durchwegs nicht gleich sein muss. An einer Stelle werden große Informationsblöcke unverändert wieder verwendet, an anderen wird Content auf Satz- oder sogar Wort-Ebene wieder verwendet. Man kann unterschiedliche Stufen von Granularität für Contenterfassung, Wiederverwendung und Ausgabe haben. Für Autoren zum Beispiel kann man Content sehr fein gliedern, um ihnen genau zu zeigen, was sie beim Schreiben verwenden können, jedoch muss man Content in jenem Grad wieder verwenden, der angemessen für die jeweilige Wiederverwendungsstrategie ist. Je feiner Content strukturiert ist, desto größer wird der Aufwand, Content zu modellieren, wieder zu verwenden und zu organisieren. Andererseits wiederum wenn Content nicht fein genug strukturiert ist, wird die Möglichkeit, Content leicht wieder zu verwenden, stark beeinträchtigt. Wenn ein Element, damit seine Bedeutung klar ist, auf umgebende Informationen angewiesen ist, sollten diese Informationen als Teil des Elements eingebunden werden. Jedoch ist Granularität auf Wort-Ebene nicht wünschenswert, denn Autoren verfassen – ungeachtet welcher Detailgrad verwendet wird – immer noch vollständige Dokumente. Wenn Granularität auf Satz-Ebene verwendet wird, schreiben Autoren nicht unzusammenhängende Sätze, die einzeln in der Datenbank gespeichert werden, um dann zu einem Dokument zusammengefügt zu werden. Im Gegenteil, Autoren schreiben ganze Dokumente, indem sie die erforderliche Granularität den

Elementen zuweisen. Granularität definiert wie ein gesamtes Dokument unterteilt, bezeichnet und für Wiederverwendung gespeichert wird. Natürlich hängt die Granularität auch vom Potential des CMS ab. Wenn man Informationen in kleinere Stücke unterteilt, wird eine große Anzahl von Einheiten erzeugt, die das CMS allesamt verarbeiten muss. Wenn nun ein Autor ein Dokument abruft, muss das System alle Einzelteile zusammensetzen und zu einem gesamten Dokument zusammenführen. Das kann sowohl das CMS als auch das Netzwerk, welches den Zugriff auf das CMS sicherstellt, sehr schnell in die Knie zwingen. Auch das Auffinden von Informationen erweist sich als schwieriger, da das CMS eine sehr große Anzahl an Einzelteilen organisieren und verarbeiten muss. Suchabfragen können unüberschaubar große Ergebnisse liefern. Andererseits wiederum kann ein gröberer Detailgrad dazuführen, dass die Wiederverwendung von Informationen von einem CMS unmöglich wird.
So, wie granular sollen nun die Informationen sein? Im Allgemeinen gibt es keine richtige Antwort. Die Auswahl, den korrekten Level an Granularität zu finden, hängt einerseits von den Informationen ab, wie man sie verwenden will, und andererseits von der Leistungsfähigkeit des eingesetzten CMS.

4.1.4 Modellierung

Nachdem man die Granularität festgelegt hat, kann man mit der Modellierung beginnen. Es sind 2 Schritte der Modellierung notwendig: Im ersten Schritt modelliert man am Level des Informationsproduktes und im zweiten erstellt man die Modellierung am Level der Informationselemente. Der erste Schritt, die Modellierung am Level des Informationsproduktes kann man vergleichen mit dem Erstellen des Grundrisses eines Hauses. Der zweite Schritt ist gleichzusetzen mit dem Bestimmen des Hausinneren – die Räume, ebenso wie die Elemente in den Räumen. Wenn man ein Haus plant, bestimmt man alle Komponenten (zum Beispiel, Räume) genauso wie die Anordnung für jeden Raum. Weiters versichert man sich, dass alle mehrfach vorkommenden Komponenten, wie zum Beispiel Türen, einheitlich strukturiert sind. Das „Produkt"-Modell eines Hauses legt fest aus welchen Räumen ein Haus besteht. Das „Element"-Modell bestimmt wie jeder Raum aufgebaut ist und welche Elemente von Raum zu Raum wieder verwendet werden. Die Modelle geben auch an, welche Elemente verbindlich sind und welche optional. Manche Elemente, zum Beispiel, sind

standardmäßig in einer Küche und andere sind nicht verbindlich. Wie auch immer, der Inhalt einer Küche wird nie derselbe sein als der eines Badezimmers.

Ein Küchenmodell könnte folgendes beinhalten:

- Fenster
- Abwaschbecken
- Kochinsel (optional)
- Schränke
- Arbeitsfläche
- Tür

Und die Elemente eines Badezimmers könnten sein:

- WC
- Fenster
- Whirlpool (optional)
- Tür

Für den Zweck dieses Vergleiches kann man die Tür und das Fenster als Elemente erachten, die wieder verwendbare, gleiche Elemente sind, ungeachtet wo sie verwendet werden. Im Folgenden werden die einzelnen Schritte der Modellierung näher betrachtet:

1. Informationsprodukt-Modelle
   Ebenso wie man ein Haus plant und welche Elemente jeder Raum beinhaltet, genauso modelliert man ein Informationsprodukt (Presseerklärung, Produktbeschreibung, Benutzerhandbuch, Prospekt, etc.). Man spezifiziert welche Komponenten es beinhaltet und welche Struktur jedes Element aufweist. Das Produkt-Modell definiert Elemente, Attribute und Metadaten von Content genauso wie die Beziehungen zwischen den Elementen. Autoren halten sich an das Modell, um Informationsprodukte einheitlich zu erstellen. Dieses Modell stellt ebenfalls Informationen bereit, welche Informationstypen im jeweiligen Informationsprodukt vorkommen. Eine Presseerklärung, zum Beispiel, könnte folgende Elemente enthalten:

- Betreff
- Datum
- Kontakt
- Hauptteil
- Website-Adresse

2. Element-Modelle

   Zusätzlich zu Produkt-Modellen müssen Element-Modelle erstellt werden. Ein Element-Modell schlüsselt das Produkt-Modell weiter auf, indem es alle Komponenten beschreibt, aus denen das Informationsprodukt zusammengesetzt ist. Das Hauptteil-Element der Presseerklärung, zum Beispiel, könnte folgende Komponenten enthalten:

   - Firmenbeschreibung (kurz)
   - Ankündigung
   - Produktbeschreibung
   - Eigenschaften
   - Leistungen
   - Angebot
   - Verfügbarkeit
   - Firmenbeschreibung (lang)

Wenn man wieder an das Beispiel des Hausplanens zurückdenkt, kann man sich vorstellen, dass jedes einzelne Informationselement ein „Container" sein kann. Das Haus ist ein Container-Element für jeden Raum, und jeder Raum ist ein Container-Element für alle Komponenten, die einen Raum ausmachen. Auf diese Art und Weise können Container wieder Container enthalten. Dasselbe ist auch auf Informationen anwendbar, denn Elemente sind Container für Informationen. Zum Beispiel besteht das Datum aus Tag, Monat und Jahr. Ein Informationsprodukt-Modell ist ein Container für seine Elemente und jedes Element kann wiederum ein Container für Subelemente sein. Abbildung 4.2 zeigt am Beispiel Presseerklärung welche Elemente komplett eigenständig und welche Elemente Container für andere Elemente sind.

Semantik	Element-Typ
Betreff	Element
Datum	Container
Tag	Element
Monat	Element
Jahr	Element
Kontakt	Container
Name	Container
Vorname	Element
Nachname	Element
Telefon	Element
Email	Element
Hauptteil	Container
Firmenbeschreibung (kurz)	Element
Ankündigung	Container
Absatz	Element
Nummerierung	Element
Aufzählung	Element
Produktbeschreibung	Element
Eigenschaften	Element
Leistungen	Element
Angebot	Element
Verfügbarkeit	Element
Firmenbeschreibung	Container
Firmenbeschreibung kurz	Element
Firmenbeschreibung lang	Element
Website-Adresse	Element

Abbildung 4.2 Elemente einer Presseerklärung
Quelle: [Rock03]

4.1.5 Komponenten der Modelle

Ein Informationsmodell beinhaltet verschiedene Komponenten, wie Grundinformation, semantische Information, Metadaten, Architekturinformation und Produktinformation für jedes einzelne Element.

Grundinformation

Die Grundinformation beschreibt die gewöhnliche Bezeichnung jedes Elements in einem Container und verwendet generische Tags oder Grundelemente. Grundelemente leiten Informationstechniker bei der Implementierung des Modells. Wenn so genannte Autorensysteme (siehe 4.2.2) verwendet werden, leiten Grundelemente Autoren bei der Auswahl der richtigen Tags für das Modell. Abbildung 4.3 zeigt die semantischen Tags und die dazugehörigen Grundelemente für das Modell der Presseerklärung. Normalerweise hat man nur eine Spalte von Grundelementen, jedoch werden hier die Grundelemente des Beispiels halber für XML und Microsoft Word angezeigt.

Semantik	Grundelement	
	XML	Word
Betreff	title	Überschrift 1
Datum	Container	
Tag	PCDATA	Standard
Monat	PCDATA	Standard
Jahr	PCDATA	Standard
Kontakt	Container	
Name	Container	
Vorname	PCDATA	Standard
Nachname	PCDATA	Standard
Telefon	para	Standard
Email	para	Standard
Hauptteil	Container	
Firmenbeschreibung (kurz)	para	Standard
Ankündigung	Container	
Absatz	para	Standard
Nummerierung	ordered list	Nummerierung
Aufzählung	unordered list	Aufzählung
Produktbeschreibung	para	Standard
Eigenschaften		
Leistungen	unordered list	
Angebot	para	Standard
Verfügbarkeit	para	Standard
Firmenbeschreibung	Container	
Firmenbeschreibung kurz	para	Standard
Firmenbeschreibung lang	para	Standard
Website-Adresse	link	Hyperlink

Abbildung 4.3 Semantische Elemente und Grundelement einer Presseerklärung
Quelle: [Rock03]

Semantische Information

Die Semantik verwendet so genannte semantische „Tags", die eine spezifische Bedeutung haben und beschreiben, was jedes Element enthält. Das Informationsproduktmodell der Presseerklärung (Abbildung 4.2) legt semantisch fest, was eine Presseerklärung beinhaltet. Anstatt allgemein zur Struktur zu verweisen, werden semantische Tags verwendet, um die Struktur zu beschreiben. Zum Beispiel das Element <Website-Adresse> bestimmt genau, dass der Content, der dieses Element beinhaltet, die Website-Adresse ist. Das Gegenteil von einem semantischen Tag ist ein generisches Tag (zum Beispiel, <Absatz>). Wenn man ein generisches Tag wie <Absatz> für die Website-Adresse verwendet, ist es nicht klar, welchen Content dieses Element enthalten soll. Die Semantik definiert klar und deutlich die Struktur der Information. Wenn man keine semantischen Tags verwenden möchte, kann man stattdessen auch Metadaten verwenden, da semantische Tags die Möglichkeit, Content wieder zu verwenden beeinträchtigen können. Das Tag <Eigenschaften> kann Informationen enthalten, die in einem anderen Informationsprodukt notwendig sind, jedoch im anderen Informationsprodukt wird für die Information eine andere Struktur verwendet. Deshalb ist wieder verwendbarer Content durch das semantische Tag nicht feststellbar.

Metadaten

Im Modell muss auch aufgezeigt werden welche Metadaten sich auf welche Elemente beziehen. Metadaten liefern Daten über Daten, oder Informationen über Informationen. Während semantische Tags beim Verfassen von Texten und Wiederverwendung helfen, bieten Metadaten Hilfe bei Suchkriterien, ähnlich wie Indexeinträge. Metadaten werden benötigt, um Content einheitlich zu kennzeichnen, damit Autoren ihn finden, wieder verwenden und verschieben können. Abhängig vom jeweiligen Autorensystem können einige Metadaten automatisch zu den Elementen hinzugefügt werden (wie es im Modell definiert ist), andere von den Autoren beim Erstellen von Content ausgewählt werden. Weitere Informationen über Metadaten und ihre Rolle bei Content Wiederverwendung finden sie unter Kapitel 4.3.2.

Architekturinformation

Die Architekturinformation liefert Details für die Wiederverwendungsmethode (siehe 3.4), ob semantische Tags oder Metadaten verwendet werden und wo Content wieder verwendet wird, bzw. wie (zum Beispiel, gesperrt [G] oder abgeleitet [A]; siehe 3.5) er wieder verwendet wird. Abbildung 4.4 stellt die Architekturinformation anhand der Presseerklärung dar.

Semantik	Architekturinformation					
	Reuse	Richtlinie	Reuse-Plan			
			Produktblatt	Prospekt	Website	E-Katalog
Betreff		Semantik				
Datum		Semantik				
Tag		Semantik				
Monat		Semantik				
Jahr		Semantik				
Kontakt		Semantik			A	A
Name		Semantik			A	A
Vorname		Semantik			A	A
Nachname		Semantik			A	A
Telefon		Semantik			A	A
Email		Semantik			A	A
Hauptteil		Semantik			A	
Firmenbeschreibung (kurz)	systematisch	Semantik	G	G	G	
Ankündigung					A	
Absatz					A	
Nummerierung					A	
Aufzählung					A	
Produktbeschreibung	systematisch	Semantik	G	G	G	G
Eigenschaften			G	G	G	G
Leistungen			G	G	G	G
Angebot					G	
Verfügbarkeit	systematisch	Semantik				
Firmenbeschreibung	systematisch	Semantik	G	G	G	
Firmenbeschreibung kurz	systematisch	Semantik	G	G	G	
Firmenbeschreibung lang	systematisch	Semantik	G	G	G	
Website-Adresse	systematisch	Semantik	G	G		

Abbildung 4.4 Architekturinformation der Presseerklärung
Quelle: [Rock03]

Produktinformation

Die Produktinformation leitet den Informationstechniker bei der Erstellung von Stylesheets oder Templates. Abbildung 4.5 zeigt die zum Modell der Presseerklärung hinzugefügte Produktinformation.

Semantik	Produktion
Betreff	
Datum	
Tag	
Monat	
Jahr	
Kontakt	
Name	
Vorname	
Nachname	
Telefon	
Email	
Hauptteil	
Firmenbeschreibung (kurz)	
Ankündigung	
Absatz	
Nummerierung	
Aufzählung	
Produktbeschreibung	
Eigenschaften	
Leistungen	
Angebot	Dieses Element sollte in Papierform kursiv angezeigt werden, online jedoch unter Anführungszeichen.
Verfügbarkeit	
Firmenbeschreibung	
Firmenbeschreibung kurz	
Firmenbeschreibung lang	
Website-Adresse	

Abbildung 4.5 Produktinformation der Presseerklärung
Quelle: [Rock03]

Abbildung 4.2 bis Abbildung 4.5 zeigt nur diverse Ausschnitte des Models, da einzelne Spalten weggelassen wurden, um nur die zur Erklärung relevanten Informationen anzuzeigen. Das komplette Model zeigt Abbildung 4.6.

Semantik	Element-Typ	Grundinformation		Architekturinformation						Produktion
		XML	Microsoft Word	Reuse	Richtlinie	Reuse-Plan				
						Produktblatt	Prospekt	Website	E-Katalog	
Betreff	Element	title	Überschrift 1		Semantik					
Datum	Container	Container			Semantik					
Tag	Element	PCDATA	Standard		Semantik					
Monat	Element	PCDATA	Standard		Semantik					
Jahr	Element	PCDATA	Standard		Semantik					
Kontakt	Container	Container			Semantik			A	A	
Name	Container	Container			Semantik			A	A	
Vorname	Element	PCDATA	Standard		Semantik			A	A	
Nachname	Element	PCDATA	Standard		Semantik			A	A	
Telefon	Element	para	Standard		Semantik			A	A	
Email	Element	para	Standard		Semantik			A	A	
Hauptteil	Container	Container			Semantik			A		
Firmenbeschreibung (kurz)	Element	para	Standard	systematisch	Semantik	G	G	G		
Ankündigung	Container	Container						A		
Absatz	Element	para	Standard					A		
Nummerierung	Element	ordered list	Nummerierung					A		
Aufzählung	Element	unordered list	Aufzählung					A		
Produktbeschreibung	Element	para	Standard	systematisch	Semantik	G	G	G	G	
Eigenschaften	Element					G	G	G	G	
Leistungen	Element	unordered list				G	G	G	G	
Angebot	Element	para	Standard					G		Dieses Element sollte in Papierform kursiv angezeigt werden, online jedoch unter Anführungszeichen.
Verfügbarkeit	Element	para	Standard	systematisch	Semantik					
Firmenbeschreibung	Container	Container		systematisch	Semantik	G	G	G		
Firmenbeschreibung kurz	Element	para	Standard	systematisch	Semantik	G	G	G		
Firmenbeschreibung lang	Element	para	Standard	systematisch	Semantik	G	G	G		
Website-Adresse	Element	link	Hyperlink	systematisch	Semantik	G	G			

Abbildung 4.6 Komplettes Modell der Presseerklärung
Quelle: [Rock03]

4.1.6 Bestehende Informationsmodelle

So wie es für Häuser bestehende Pläne zu kaufen gibt, gibt es auch Informationsmodelle zu kaufen, bzw. bestehende Informationsmodelle, die von anderen entwickelt und für jeden frei zugänglich gemacht wurden. Diese Modelle sind definiert als SGML/XML DTDs, und sollten lediglich als Starthilfe für das eigene Modell dienen. Durch die zunehmende Popularität von XML (siehe 4.2.3) werden fast täglich neue Modelle veröffentlicht. Einige bieten Lösungen für sehr spezifische Anforderungen, wie zum Beispiel das Finanzverlagswesen, andere bieten ein breiteres Spektrum und können für unterschiedliche Problemstellungen herangezogen werden.

Im Folgenden werden drei verfügbare Modelle erwähnt:

- **DocBook**

  DocBook ist ein sehr bekanntes Informationsmodell für Softwaredokumentation. Es wurde ursprünglich als eine „SGML markup language“ entwickelt, um technische Handbücher für UNIX Computersysteme gemeinsam zu nutzen und letztendlich zu erstellen. Es ist ein sehr robustes Modell, welches so ziemlich jedes Modell eines Systemsoftware-Handbuches anpassen kann. Augrund der langen Entwicklungsphase (über 10 Jahre) ist es ein sehr stabiles, jedoch auch sehr komplexes, nicht gerade leicht verständliches Modell.

- **DITA**

  DITA (Darwin Information Typing Architecture) basiert auf XML und bietet eine durchgehende Architektur, um technische Informationen zu verfassen. Die Architektur und daraus resultierende DTD wurden von einem Team von IBM, Lotus und Tivoli entwickelt. Ihr Ziel war es, eine Architektur zu entwickeln, die Informationsaustausch und -wiederverwendung unterstützt. DITA ist jedoch nicht wirklich ein direkt implementiertes Modell, stattdessen definiert es durch Grundmodelle, wie man sein eigenes Modell bilden kann. DITA ist ein sehr mächtiges Model, jedoch aufgrund des relativ jungen Bestehens gibt es sehr wenige Implementierungen, von denen man lernen bzw. welche man adaptieren könnte.

- **SCORM**

  Das Shareable Content Object Reference Model (SCORM) ist XML-basiert und ein Standard von ADL (Advanced Distributed Learning), einer Initiative des amerikanischen Verteidigungsministeriums und des Ministeriums für Forschung und Technologie. Es hat das Ziel, Lerninhalte wieder verwendbar, kompatibel mit verschiedenen Lernumgebungen, dauerhaft und unabhängig von Betriebssystemen nutzbar und jederzeit zugänglich zu machen. Durch eine derzeit große Menge an oft frei verfügbaren Autorensystemen wird das Erstellen von SCORM-konformen Inhalten unterstützt.

Nähere Informationen über diese Modelle bzw. über weitere veröffentlichte Modelle findet man auf der Website des World Wide Web Consortiums (W3C) unter `http://www.w3.org`.

Wenn man plant, zu XML und zu strukturiertem Content zu wechseln, können diese fertigen Modelle die optimale Lösung sein. Will man jedoch sehr aufwendig Content wieder verwenden und ist die Struktur des Contents der Informationsstruktur im Unternehmen angepasst, werden solche Modelle wahrscheinlich nicht geeignet sein. Hierbei wird man nicht darum herumkommen, ein bestehendes Modell zu adaptieren bzw. zur Gänze ein eigenes Modell zu erstellen.

4.1.7 Wie werden diese Modelle eingesetzt?

Wenn Modelle einmal erstellt wurden, müssen sie im gesamten Unternehmen eingeführt werden, damit Autoren und Rezensenten sie verwenden können, um Content zu erstellen, zu editieren und durchzusehen. Autoren verwenden Informationsmodelle, um festzustellen, welche Informationen zu welchem Informationsprodukt gehören und wie jedes Element aufgebaut ist. Zum Beispiel können sie aus Modellen herauslesen, dass ein Informationsprodukt ein Überblick-Element benötigt, und dass dieses Element eine bestimmte Struktur hat. Weiters können sie Regeln und Hinweise erlangen (Schreibnotizen), wie manche Komponenten geschrieben werden. Autoren können Elemente vom CMS auswählen, welche laut Modell eingefügt werden müssen. Rezensenten verwenden diese Modelle, um Entwürfe der Autoren zu begutachten. Sie vergleichen den Entwurf mit dem Modell, um sicherzustellen, dass alle notwendigen Elemente vorhanden sind. Informationstechniker verwenden die Modelle, um so genannte „Authoring Templates“ oder DTDs zu erstellen bzw. das CMS zu implementieren.
Im nächsten Abschnitt wird die Umsetzung und Implementierung dieser Informationsmodelle aufgrund verschiedener möglichen Varianten erläutert.

4.2 Umsetzen von Informationsmodellen

Man kann das erstellte Model auf verschiedene Weisen, abhängig von Autoren-Anforderungen, eingesetzten Technologien und gewünschten Ergebnissen umsetzen. Im Folgenden werden unterschiedliche Arten der Implementierung vorgestellt.

4.2.1 Autorenformulare

Autorenformulare (authoring forms) sind HTML-Formulare, welche Autoren bei der Eingabe von strukturiertem Content leiten. Sie geben Content in Formularfelder im Browser ein. Authoring forms können in Verbindung mit XML und DTDs einerseits als Alternative zu DTDs und andererseits als Interface zum CMS verwendet werden. Solche formularbasierten Autorenwerkzeuge sind typischerweise oft Teil eines CMS. Das CMS bietet Möglichkeiten, HTML-Formulare mit individuellen Feldern zu erstellen, um Contentmodell-Elemente genau abzubilden. Diese Formulare werden oft nur für das Erstellen gewisser Arten von Content wie Pläne, Eigenschaften, Richtlinien und Methoden, hauptsächlich kleine Online-Inhalte verwendet.

Authoring forms haben folgende Vorteile gegenüber vollständiger Autorensysteme:

- Formulare ermöglichen gemeinsames Erstellen von Inhalten mit Zugriff von außen. Jeder, der über einen Browser verfügt und Zugriff zum Intranet, Internet oder Extranet hat, kann Content über Formulare eingeben.
- Formulare verstecken die Komplexität, Inhalte in XML einzugeben vor den Autoren. Autoren müssen keine Styles oder Tags zuweisen, sie müssen lediglich Content in die Felder, die ihnen das Formular zur Verfügung stellt, eingeben. Wenn das Formular dann zum Server geschickt wird, werden alle Daten entsprechend konvertiert.

Natürlich gibt es auch Einschränkungen:

- Formulare benötigen IT-Support, um sie zu erstellen und die Daten aus den Feldern zu extrahieren bzw. zu konvertieren.
- Formulare sind bei großen Informationsproduktmodellen nicht zweckmäßig.
- Formulare sind sehr unflexibel, da bei Änderungen, Änderungen im Code notwendig sind.
- Formulare unterstützen keine feine Granularität bei der Wiederverwendung.

4.2.2 Autorensysteme

Autorensysteme, oder auch Strukturtemplates (structural templates) genannt, sind aufbereitete Templates, die strukturierte Namen verwenden, um die Strukturelemente eines Dokumentes abzubilden. Sie verwenden semantische Namen und finden hauptsächlich bei Textverarbeitungs- und Desktoppublishingsystemen Anwendung. Strukturtemplates besitzen nicht dieselbe Stärke wie DTDs oder Formulare, weil die Tools, die diese Templates verwenden, nicht kontrollieren können, wann Tags verwendet werden und wann nicht. Ein XML-Editor, zum Beispiel, kann herausfiltern, welche Tags ein Autor verwenden darf, und daraus resultiert eine DTD. Wenn eine DTD eine Reihenfolge an Tags definiert (Titel, gefolgt von einer Einleitung, gefolgt von einer Kurzbeschreibung, …), kann ein Autorensystem lediglich sicherstellen, dass alle Tags verwendet werden. Es gibt keinen Mechanismus, der Autoren daran hindert, die Struktur oder Tag-Namen zu verändern oder sogar neue Tags zu erstellen.

Structural Templates können sehr gut für Textbausteine oder Hinweise zu gebräuchlichen Informationen verwendet werden.

4.2.3 XML

Ähnlich wie HTML (**H**yper**t**ext **M**arkup **L**anguage) ist auch XML (e**X**tensible **M**arkup **L**anguage) eine Auszeichnungssprache. Anders jedoch als bei HTML, wo definierte (generische) „Tags" vorgegeben werden, gibt XML nur die Standards vor, auf jenen diese Tags beruhen. Das heißt mit Hilfe von XML werden eigene (semantische) Tag-Namen definiert, um die gewünschten Informationen auszeichnen zu können. Idealerweise ergeben sich diese Namen aus den semantischen Tags des zugrundeliegenden Informationsmodells. Durch das Definieren eigener Tag-Namen ergeben sich folgende Vorteile:

- Tag-Namen haben eine spezielle Bedeutung. Sie können so spezifisch und präzise definiert werden, dass Autoren ihren Zweck und ihre Bedeutung nicht erraten müssen, sondern sie wissen genau wofür sie verwendet werden.
- Die Namen spiegeln den Content wider. Die Tag-Namen bestimmen eindeutig was sie enthalten.

- Tag-Namen haben nichts mit Formatierung zu tun. Die Formatierung kann später definiert werden, wann der Zweck des Dokumentes genau bekannt ist.
- Je nach Bedarf kann man wenige oder viele Tags verwenden. HTML hat eine fixe Anzahl an vordefinierten Tags. Dabei ist es sehr schwierig, Autoren davon abzuhalten, die Tags zu verwenden, die sie nicht verwenden sollen.

Die erlaubten XML-Strukturen werden durch DTDs (**D**ocument **T**ype **D**efinition) exakt festgelegt und sind eine formale Definition der XML Elemente, die ein spezieller Typ eines Dokumentes beinhalten kann. Sie unterstützen einen stark strukturierten Weg, um Informationsmodelle besser umsetzen zu können. Eine DTD definiert die Namen der Elemente, ihre Beziehungen und ihre Vorkommen. XML benötigt nicht unbedingt eine DTD, jedoch will man Content wieder verwenden, sind DTDs zwingend erforderlich, denn der Schlüssel zu effektiver Wiederverwendung ist Konsistenz, die durch DTDs sichergestellt wird. Die vermehrte Nutzung von XML hat dazu geführt, dass eine Alternative zu DTDs entwickelt wurde: XML Schemas. XML Schemas sind im Grunde dasselbe wie DTDs – sie definieren Struktur. Jedoch haben XML Schemas den Vorteil, dass sie den Elementen auch Datentypen zuweisen können. Zum Beispiel, wenn ein Element ein Datum darstellt, kann genau definiert werden in welchem Datumsformat der Content sein soll.

4.3 Identifikation von Content

Nachdem man ein Informationsmodell entwickelt hat, will man alle Contentteile leicht finden, wieder verwenden oder verändern. Deswegen müssen alle Contentelemente mit einem eindeutigen Namen, einer so genannten „semantischen Identität“ versehen werden. Im Folgenden werden zwei Möglichkeiten näher erläutert, wie man Textelementen eine semantische Identität zuweisen kann.

4.3.1 Stylesheets

Stylesheets werden je nach eingesetzter Technologie unterschiedlich verwendet. Wenn traditionelle Autorensysteme zum Einsatz kommen, werden Stylesheets einerseits für das Layout des Dokumentes im Editor und andererseits für das Ausgabe-Layout verwendet. Es ist sehr wichtig, dass man den Autoren die Möglichkeit geben

kann, die Anzeige der Informationen im Editor selbst zu definieren. Dadurch ist die Editieroberfläche auf seine Bedürfnisse angepasst, und ein effizienteres Arbeiten ist möglich. Auf jeden Fall sollten die Autoren unterrichtet werden, dass ihre Anpassungen keine Auswirkungen auf die Ausgabe haben. Nachdem die Contenterfassung abgeschlossen ist, veröffentlicht der Content Management Server die Seite mit dem richtigen Ausgabe-Stylesheet. Während Autoren ihre Styles selbst verändern können, sollten Ausgabetemplates gesperrt sein, um sie konsistent zu halten.

Stylesheets, die für XML erstellt werden (XSL Stylesheets), dienen nicht nur für die Definition des Layouts, sondern bieten auch weitere Möglichkeiten wie:

- Sortierung
- Unterstützung von standardisierten Textelementen
- Verbergen von Text
- Wiederholen bzw. Umgestalten von Text

Diese Funktionen sind sehr mächtig, da diese Styles einfach erstellt werden können und nicht kompiliert werden müssen (im Gegensatz zu VisualBasic oder Javascript). Der Parser kann die Änderungen direkt auf ein Dokument anwenden.

4.3.2 Metadaten

Ohne effektive Metadaten können Autoren nur sehr schwer die richtigen Informationen finden, geschweige denn wieder verwenden. Metadaten sind Daten über Daten. Sie werden verwendet, um Prozesse, Regeln, Strukturen und Daten zu beschreiben und um näher erläuternde Informationen zu Daten hinzuzufügen. Es gibt zwei Arten von Metadaten:

- Klassifizierende Metadaten

  Diese Metadaten klassifizieren bzw. katalogisieren Dokumente. Sie werden verwendet, um Content abzurufen und wieder zu finden.
- Element Metadaten

  Element Metadaten kennzeichnen Content am Level der Elemente, die im Informationsmodell definiert sind. Autoren werden durch diese Metadaten im Contenterstellungsprozess unterstützt. Es gibt drei Haupttypen dieser Metadaten:

1. Metadaten für Wiederverwendung werden verwendet um Contentkomponenten, die in verschiedenen Bereichen wieder verwendet werden können, zu identifizieren. Zum Beispiel, wenn eine Produktübersicht für ein Produkt ABC schon existiert, können Metadaten wie „Contenttyp=Überblick, Produkt=ABC“ verwendet werden, um den richtigen Content zur Wiederverwendung finden zu können.
2. Metadaten für Wiederfindung werden verwendet um Content abzurufen. z. B.: Autor, Erstellungsdatum, Änderungsdatum, Schlüsselwörter, etc.
3. Metadaten für Statusinformationen werden verwendet um den Status des Contents in einem Workflowsystem festzustellen. z.B.: Überprüft, Freigegeben, Veröffentlicht, Archiviert, etc.

Während der Umsetzung des Informationsmodells muss überlegt werden, wo Metadaten gespeichert werden. Der Speicherort hängt vom eingesetzten CMS und dem verwendeten Datenformat ab. Normalerweise speichert ein CMS Metadaten in einer Datenbanktabelle.

In XML können Metadaten in Elementen oder Attributen gespeichert werden. Die meisten Schnittstellen zwischen XML Autorensystemen und Content Management Systemen beinhalten Funktionalitäten, um Metadaten aus der XML Datei zu extrahieren und in die Metadatenfelder des CMS einzufügen. Wünschenswert wäre natürlich, wenn dieser Prozess bidirektional möglich ist. Wenn, zum Beispiel, Metadaten in der Quelldatei aktualisiert wurden, sollten die Änderungen in der CMS Metadaten Datenbank gespeichert werden, wenn die Datei „eingecheckt“ wird. Wenn Metadaten im CMS verändert werden, sollten diese Änderungen auch in der Quelldatei nachgezogen werden.

Für traditionelle Autorensysteme werden Metadaten im CMS gespeichert.

4.4 Bedeutung von XML für Content Wiederverwendung

XML hat einige Vorteile gegenüber traditioneller Autorensysteme, um Content wieder zu verwenden. Im Folgenden werden die Eigenschaften, die Content Wiederverwendung fördern genauer erklärt.

4.4.1 Strukturierter Content

Üblicherweise besitzen Autoren ein großes Verständnis für strukturierten Content. Sie wissen, zum Beispiel, dass Bücher aus einem Titel, mehreren Kapiteln und mehreren Anhängen bestehen. Sie erkennen auch wiederkehrende Strukturen auf einem niedrigeren Level: Kapitel haben eine Überschrift, einen Überblick, mehrere Abschnitte, die den Hauptteil bilden und eine Zusammenfassung. Jedoch wenn man gleichartige Informationsprodukte näher betrachtet, sieht man, dass die Strukturen von einem zu einem anderen Produkt nicht konsistent sind. Strukturen unterscheiden sich von einem Autor zu einem anderen, von einer Abteilung zu einer anderen und auch von einem Geschäftsbereich zu einem anderen. Es unterscheiden sich oft sogar Informationen, die von einem einzigen Autor verfasst wurden, über eine gewisse Zeitspanne. Das stellt natürlich ein sehr großes Problem für Content Wiederverwendung dar.

In XML wird die Struktur durch DTDs oder Schemas definiert. Eine DTD ist sehr spezifisch, da sie alle Elemente (XML Tags) definiert, die in einem Dokument verwendet werden dürfen. Sie definiert weiters die Beziehung dieser Elemente zu anderen Elementen. Man kann auch die Hierarchie, die Anordnung und die Anzahl der Elemente festsetzen. Eine DTD kann unglaublich nützlich für die Contenterstellung sein. Autoren beschäftigen sich oft länger damit die richtige Struktur für ein Dokument zu finden, als mit dem Verfassen des Dokumentes selbst. Durch eine DTD ist die erforderliche Struktur vorgegeben. Diese Konsistenz ist auch sehr hilfreich für die Informationskonsumenten. Sie lernen, wo man Informationen schnell und effizient finden kann.

Für eine strukturelle Konsistenz ist eine definierte Struktur in einer DTD die halbe Lösung. Die andere Hälfte wird durch spezielle Validierungstools abgedeckt, die eine DTD lesen und die darin definierten Strukturregeln erzwingen können. Stellt man Autoren solche Tools und eine DTD bereit, kann man sicher sein, dass Informationsprodukte eine konsistente Struktur besitzen.

4.4.2 Trennung von Content und Format

Traditionelle Autorensysteme, wie zum Beispiel Textverarbeitungsprogramme, haben einen großen Nachteil: Sie wurden entwickelt, Autoren dabei zu unterstützen, dass Dokumente gut aussehen; das heißt ihr Hauptaugenmerk liegt auf der Formatierung. Jedoch aus Sicht der Wiederverwendung ist das nicht unbedingt ein Vorteil. Textverarbeitungsprogramme erlauben Autoren ihre Formatierung selbst zu definieren, die bestehende Formatierung zu ändern bzw. Formatierungen generell zu ignorieren. Es gibt keine Funktionalität, die eine festgesetzte Formatierung konsistent hält. Der gesamte Content und auch alle Formateinstellungen werden in ein und derselben Datei gespeichert, was natürlich auch unnötig zu sehr großen Dateien führt. Um den Content nun wieder verwendbar zu machen, müsste man einen Weg finden, um die Formatierung aus der Datei entfernen und den Content unabhängig davon machen zu können. Dies ist natürlich nicht möglich bzw. mit sehr viel Handarbeit verbunden.

XML hat einen wesentlichen Vorteil gegenüber den Textverarbeitungsprogrammen. XML stammt von der Grundidee ab, Dokumente über Systeme und Applikationen transportierbar zu machen. Jeder Befürworter von Bezeichnungssprachen weiß, dass jede eingebettete Formatierung und das Binärdateiformat Hindernisse für eine Plattform unabhängige Transportabilität sind. Die Lösung ist, Content von Format zu trennen. XML richtet sich auf die Struktur eines Dokumentes und nicht auf die Darstellung. Darstellungsinformationen werden in separaten Dateien gehalten, die mit dem Dokument verbunden werden, wenn es veröffentlicht wird.

Die Trennung von Content und Format bringt eine erhebliche Flexibilität mit sich. Ein Hinweiselement, zum Beispiel, könnte folgendermaßen für die Ausgabe formatiert sein:

HINWEIS:	Wenn Sie Ihren Benutzernamen oder ihr Passwort vergessen haben, kontaktieren Sie bitte Ihren Systemadministrator.

Abbildung 4.7 Einfach formatierte Mitteilung
Quelle: [Rock03]

Das Wort „HINWEIS“ ist nicht Bestandteil des Contents; es wurde durch das Stylesheet hinzugefügt. Dieses Schlüsselwort kann mit Hilfe des Stylesheets leicht durch ein Symbol ersetzt werden:



Wenn Sie Ihren Benutzernamen oder ihr Passwort vergessen haben, kontaktieren Sie bitte Ihren Systemadministrator.

Abbildung 4.8 Mitteilung mit Symbol
Quelle: [Rock03]

4.4.3 Eingebettete Metadaten

Wie schon erwähnt ist HTML als Bezeichnungssprache definiert, während XML eine Menge von Regeln ist, um Tags zu erstellen. Was bedeutet das nun, wenn man genauer hinsieht? Die Dateien beinhalten dieselbe Information, allerdings bieten die Tag-Namen für Autoren zusätzliche Details über die Information. Die Tag-Namen sind Metadaten.

Für Begebenheiten, wo zusätzliche Informationen notwendig sind, um Content zu beschreiben, können Attribute verwendet werden, um weitere Metadaten zu definieren. Zum Beispiel können Metadaten in einer Wiederverwendungsumgebung dazu verwendet werden, um zu kennzeichnen, wer die beabsichtigte Zielgruppe für ein bestimmtes Stück an Information ist. Nachstehendes Beispiel zeigt anhand von einer Prozedur zum Einloggen wie solche Attribute verwendet werden. Zum Tag „Schritt“ wurde ein Attribut (`Produkt="erweitert"`) hinzugefügt. Dieses Attribut fungiert im Sinne von Metadaten, indem es angibt, dass dieser Schritt nur für die erweiterte Version des Produktes verwendbar ist.

```
<Prozedur>
   <Titel>AccSoft-Anmeldung</Titel>
   <Absatz> Wenn sie das erste Mal auf eine Komponente von AccSoft
      drücken, müssen Sie sich am System anmelden.</Absatz>
   <Einführung>Um sich anzumelden, sind folgende Schritte
      nötig.</Einführung>
   <Prozedur_Schritte>
      <Schritt>Klicken Sie die Applikation doppelt an.</Schritt>
      <Schritt>Geben Sie Ihren Benutzernamen im Namensfeld
         ein.</Schritt>
      <Schritt>Geben Sie Ihr Passwort ins Passwortfeld
         ein.</Schritt>
      <Schritt Produkt="erweitert">Wählen Sie den Kunden aus,
         den Sie aktualisieren wollen.</Schritt>
      <Schritt>Klicken Sie den OK Button an.</Schritt>
   </Prozedur_Schritte>
   <Übung>Melden Sie sich mit Ihrem Benutzernamen und mit Ihrem
      Passwort, das Sie von Ihrem Vortragenden erhalten haben,bei
      der Übungsdatenbank an.</Übung>
   <Mitteilung>Wenn Sie Ihren Benutzernamen oder ihr Passwort
      vergessen haben, kontaktieren Sie bitte Ihren Systemadmi-
      nistrator.</Mitteilung>
   <Warnung>Diese Datenbank enthält vertrauliche Informationen
      über unsere Kunden. Stellen Sie sicher, dass kein anderer
      ihre Berechtigung benutzt.</Warnung>
</Prozedur>
```

Abbildung 4.9 XML mit Attribut-Metadaten
Quelle: [Rock03]

Bei traditionellen Content Management Systemen können Autoren durch ein Auswahlfenster Metadaten zu einer Datei hinzufügen. Die besten CMS Produkte beinhalten auch Suchfunktionen über Metadaten. Jedoch sind diese Metadaten nicht Bestandteile der Dateien, sondern sie werden nur mit den Dateien verbunden. Wenn man also eine Datei auf einen anderen Ort verschiebt, werden die Metadaten nicht mittransferiert, da diese Bestandteil des CMS sind. In XML wandern alle Metadaten mit der Datei mit. Sie werden beim Aktualisieren der Datei hinzugefügt. Die Metadaten bleiben Teile des Contents und können leicht gesucht werden.

4.4.4 Datenbankorientierung

Der Prozess, die Struktur der Informationen und die daraus resultierende DTD zu bestimmen, ist sehr ähnlich zur Analyse, die ein Entwickler durchführen muss, der ein Datenbankdesign erstellt. Datenbankdesigner sind nicht an den Daten selbst interessiert, sondern an den Datentypen, der Hierarchie der Informationen und der Beziehungen zwischen den Elementen.

Eine ähnliche Methode wird verwendet, wenn man eine XML-Struktur definiert. Das Ergebnis ist ein strukturelles Format, das leichter in der Datenbank in Form von ein-

zelnen Elementen gespeichert werden kann als ein ganzes Dokument. Diese Elemente können dann in jeder beliebigen Reihenfolge ausgelesen und zusammengesetzt werden.

4.4.5 XSL Stylesheets

Durch die Trennung von Content und Format muss man sich speziell mit der Präsentation der Informationen auseinandersetzen. XML selbst bietet keine Möglichkeiten die Informationen in angemessener Form darzustellen. Die Technologie, die zum Einsatz kommt, wenn man XML darstellen will, ist XSL (e**X**tensible **S**tylesheet **L**anguage). Im Gegensatz zu traditionellen Stylesheets, welche nur Formatierungsmöglichkeiten bieten, ist XSL ein sehr mächtiges Werkzeug, um XML Dokumente zu transformieren und zu formatieren.

XSL ist selbst eine XML Auszeichnungssprache und kann als solche Content für Onlinedarstellungen und Dokumentenausgabe formatieren, Konstanten und Grafiken hinzufügen (wie Abbildung 4.8), Content filtern und Text sortieren und reorganisieren. XSL besteht derzeit aus drei großen Teilen:

- XPath

  XPath ist ein Mechanismus, um spezifische Elemente in einem XML Dokument zu identifizieren und zu formatieren, und ermöglicht das Hinzufügen von Logik zur Formatierung. In XSL Stylesheets kann man Elemente identifizieren und spezielle Formatierungen oder Transformationen auf Elemente anwenden, wie zum Beispiel, dass vor jedem Kapitel eine Überschrift stehen muss oder was der erste Absatz in einem Abschnitt ist, etc.

- XSLT (**T**ransformations)

  XSL ermöglicht nicht nur einfach das Formatieren von Informationen in einem Dokument, sondern bietet auch die Möglichkeit, Informationen in andere zu transformieren. So kann man Informationen neu ordnen, herausfiltern, oder sogar Informationen neu hinzufügen. Hier fügt sich XSLT ein. XSLT macht es möglich, XML Dokumente in andere Auszeichnungssprachen umzuformen. Hauptsächlich wird XSLT dazu verwendet, um XML Dokumente in HTML zu transformieren, um sie im Web anzeigen zu können. Jedoch wird XSLT auch genutzt um Informationen von XML in Anzeigesprachen für

PDAs (**P**ortable **D**igital **A**ssistant, Pocket PC) oder Mobiltelefone zu konvertieren.

- XSL-FO (**F**ormatting **O**bjects)
  Trotz des enormen Wachstums des Internets ist Papier immer noch ein notwendiges Ausgabeformat. XSL-FO ist für diesen Grund entwickelt worden. XSL-FO besitzt Stylesheet-Fähigkeiten, um XML zu Papierformaten wie PDF (**P**ortable **D**ocument **F**ormat) zu konvertieren. Es besitzt alle erforderlichen Formatierungsmöglichkeiten, wie Seitenlayout, Kopfzeile, Fußzeile, gerade und ungerade Seiten, Hochformat und Querformat, etc.

Die Flexibilität von XSL und ihrer Bestandteile ist extrem wertvoll für die Veröffentlichung und Anzeige von Informationen. Im Gegensatz zu traditionellen Stylesheets, die ein einziges Stylesheet mit einem Dokument verbinden können, können beliebig viele Stylesheets für ein einzelnes XML Dokument erstellt werden. Wenn dann die Information bereit zum Veröffentlichen ist, können alle Stylesheets gleichzeitig verarbeitet werden und alle Ausgabeformate werden zur selben Zeit generiert.

4.4.6 Personalisierung

Personalisierung ist sehr beliebt für die Anzeige von Informationen im Web. Personalisierung - einfach gesagt - ist Information, die manipuliert wird, um gewisse Benutzerbedürfnisse abzudecken. Diese Manipulation kann entweder vom Benutzer selbst definiert oder von der Software verwaltet werden. Mithilfe von XML können Dokumente aufgeteilt und die einzelnen Elemente separat in der Datenbank oder im CMS gespeichert werden. Diese Elemente können dann je nach Benutzeranforderungen ausgegeben werden.

Kapitel 5

Content Wiederverwendung in verteilten CMS

Eigentlich wurden Content Management Systeme in den Anfängen so konzipiert, dass sie aus einer Ansammlung von zentralisierten Applikationen bestehen und die Daten zentral abgespeichert werden. Wenn nun verschiedenste Systeme in einem Unternehmen eingesetzt wurden, wurde der größte Teil von Content in den unterschiedlichsten zueinander nicht kompatiblen Formaten, abhängig von den Tools, die diesen Content generierten, gespeichert. Wenn nun eine Abteilung ihre Informationen an andere Abteilungen weitergeben wollte, war es oft besser, diese Informationen auszudrucken, den Ausdruck weiterzugeben und diesen Content dann neu zu erstellen. Sogar wenn die Daten in strukturierten Datenbanken abgelegt oder programmatisch konvertiert wurden, mussten sie aufgrund von semantischen Unterschieden in den Datendefinitionen und den darunter liegenden Datenbankmodellen händisch neu organisiert und aufbereitet werden, damit sie in anderen Systemen verwendet werden konnten. Abhilfe schafften dann Entwicklungen, diese allein stehenden und nicht integrierten Systeme mit Systemen, die mit einander „sprechen“ können, abzulösen. Der erste Schritt in diese Richtung wurde unternommen, indem man Application Programming Inerfaces (API) für diese Systeme entwickelte und freigab. Somit gab es nun Schnittstellen, über die man Daten zwischen den Systemen austauschen konnte. Der nächste Schritt, und auch der wichtigste im Hinblick auf Datenaustausch wurde unternommen, indem man strukturierte Datendefinitionssprachen

wie SGML oder XML in die Systeme integrierte. Dadurch ist es nun möglich, Daten über ein definiertes Format unter den verschiedensten, sogar plattformunabhängigen Systemen auszutauschen. *[Wald04]*

Damit Content in verteilten Content Management Systemen wieder verwendet werden kann, muss er flexibel und wieder verwendbar gestaltet sein, das heißt er muss modular aufgebaut und mit möglichst vielen Metainformationen versehen sein. Die Entwicklung eines solchen Systems beinhaltet sehr viele Herausforderungen, die im Folgenden näher erläutert werden.

5.1 Einheitliche Definition und Struktur des Contents

Wenn Daten über verteilte Systeme ausgetauscht werden, ist es notwendig, dass ähnliche oder sogar identische Datenmodelle und Formate verwendet werden, um die Daten ohne große Änderungen verschieben zu können. Deswegen ist es von Vorteil, wenn man bestehende Modelle wie zum Beispiel DocBook (siehe 4.1.6) einsetzt und gegebenenfalls an unternehmensspezifische Informationsstrukturen anpasst. Die Vorteile für die Autoren liegen auf der Hand: *[Hack02]*

- Alle Autoren fügen denselben Content zu einem Thema ein.
- Autoren sind damit vertraut, dass sie die wichtigen Informationen niederschreiben und irrelevante Informationen nicht verwenden.
- Erfahrene und vor allem unerfahrene Autoren erstellen Content viel schneller.
- Neue Autoren wissen, welche Informationen von Ihnen gewünscht sind, wenn sie für ein bestimmtes Thema schreiben.
- Chefredakteure und Rezensenten wissen genau, welche Informationen vollständig und richtig sind.

5.2 Erweiterte Metadaten

Zu allen Contentteilen müssen Metadaten hinzugefügt werden, um Suchmechanismen bereitstellen und Statusinformationen erhalten zu können. Hierbei unterscheidet man die vom System generierten und die vom Benutzer hinzugefügten Metadaten. Vom System generierte Metadaten werden beim Anlegen, beim Abspeichern oder

beim Ändern automatisch hinzugefügt und sind zum Beispiel ID, Version, erstellt von, erstellt am, geändert von, geändert am, freigegeben von, freigegeben am, Contenttyp, Status (checked in/out), etc. Weiters gibt es Metadaten, die nur von Administratoren geändert werden können und in direktem Zusammenhang mit dem Informationsmodell stehen. So kann zum Beispiel für jeden Contentteil die Wiederverwendungsvariante (gesperrt oder abgeleitet, siehe 3.5) festgelegt werden oder die Benutzer bzw. Benutzergruppen konfiguriert werden, die diesen Contentteil wieder verwenden dürfen oder nicht. Metadaten wie Schlüsselwörter, Kategorie, Sprache, etc. müssen vom Contentersteller beim Erstellungsprozess hinzugefügt werden.

5.3 Beziehungen und Abhängigkeiten

Wenn Content wieder verwendet wird, müssen die Abhängigkeiten untereinander festgehalten werden. Das heißt, wenn ein Contentteil in ein Dokument eingebunden (wieder verwendet) wird, muss auf beiden Seiten die „Verlinkung“ vermerkt werden. Die Verlinkung beim Child-Element dient zum Anzeigen des gewünschten Inhalts und könnte unter Verwendung von XML wie folgt aussehen:

```
<task>
<title>Preloading and centering the redraw carriage</title>
<taskbody>
<prereq>Before you begin this procedure, the redraw carriage must
be installed in the press and the upper cam followers otated to
their highest position. Make sure that the way strips and way
followers are clean and smooth. If they are not remove, rework,
and install the way strips and way followers as explained in
<cite xlink:form="simple" href="installwfollow.xml">Removing,
installing, and setting up the way followers</cite>.
</prereq>
</task>
```

Abbildung 5.1 Verknüpftes XML-Dokument
Quelle: [Hack02]

Die Verlinkung beim Parent-Element dient zur Benachrichtigung der Autoren, die dieses Element wieder verwendet haben, bei Änderungen. Im Folgenden werden die verschiedenen Arten, wie Content untereinander in Verbindung stehen kann, erläutert.

5.3.1 Parent-Child-Beziehung

Content wird – wie er ist – wieder verwendet und kann nicht verändert werden (Gesperrte Wiederverwendung).

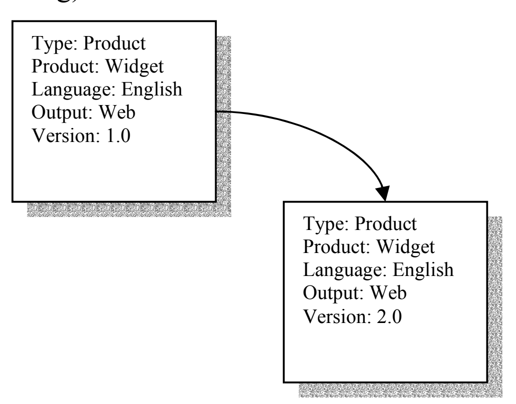


Abbildung 5.2 Parent-Child-Beziehung
Quelle: [Rock03]

5.3.2 Geschwister-Beziehung

Wenn man abgeleiteten Content für ein Element erstellt, ist das abgeleitete Element ein Child-Element des Parent-Elements. Mehrere „Kinder“ desselben Parent-Elements sind „Geschwister“.

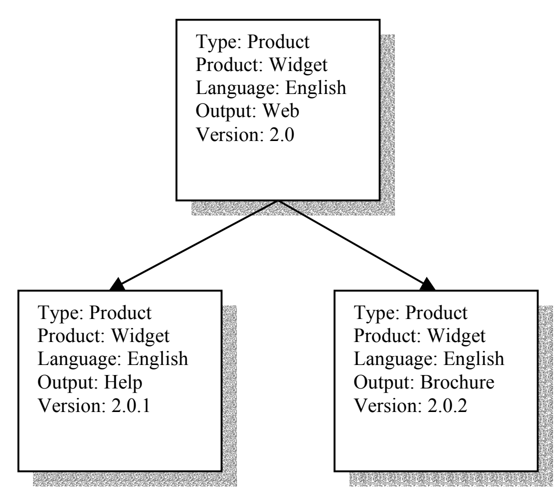


Abbildung 5.3 Geschwister-Beziehung
Quelle: [Rock03]

5.3.3 Komplexe Beziehung

Wenn Elemente auch noch übersetzt werden, so wird noch eine dritte Ebene eingeführt. Die übersetzten Elemente stehen wiederum zueinander in einer Parent-Child-Beziehung und sind somit „Enkelkinder“ vom Stammelement.

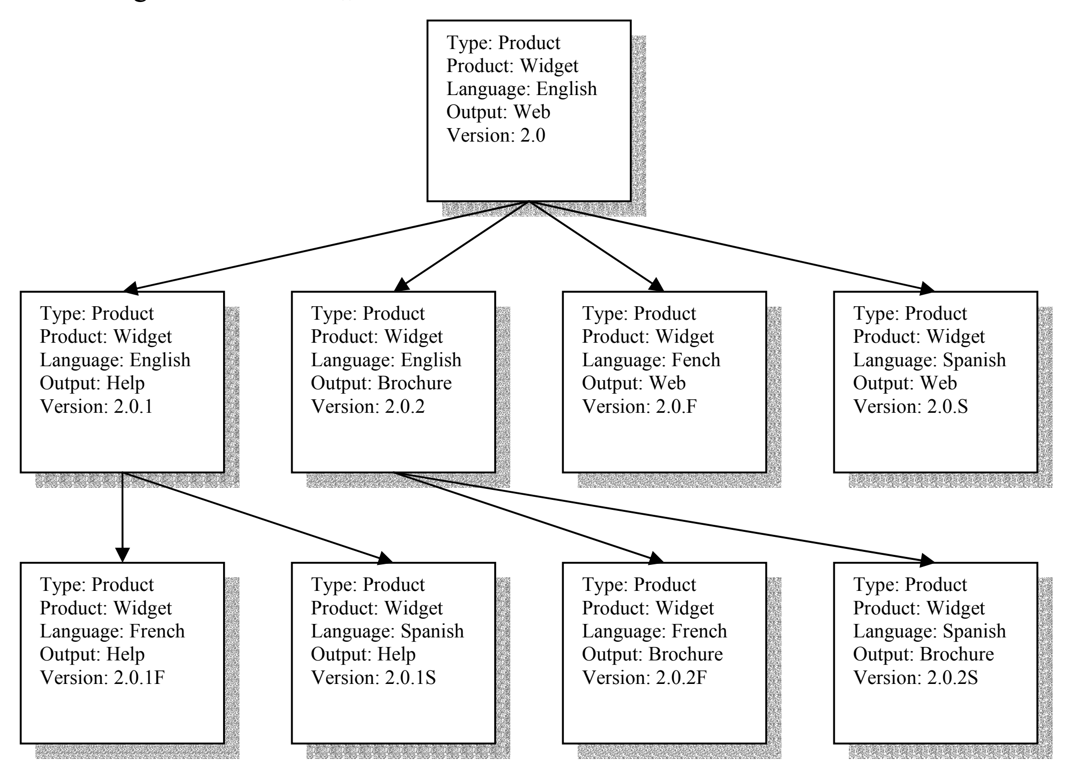


Abbildung 5.4 Komplexe Beziehung
Quelle: [Rock03]

5.4 Suchfunktionen

Das Suchen und Finden von relevanten Informationen ist das Wichtigste bei der Content Wiederverwendung in verteilten Content Management Systemen. Ohne geeignete Möglichkeiten, Contentelemente schnell und leicht zu finden, ist Content Wiederverwendung nicht möglich. Bevor Autoren sich einige Zeit mit der Suche herumschlagen, schreiben sie die gewünschten Teile neu, und Content Wiederverwendung wird nicht eingesetzt. Damit die Suche effizient ist, sollte wenn möglich keine Volltextsuche über alle Inhalte durchgeführt werden, da der gesamte Content durchsucht werden muss und dies ein sehr zeitaufwendiger Prozess ist. Wenn der gesamte Content strukturiert und mit Metadaten versehen ist, sollte die Suche über

die Metadaten durchgeführt werden. Das setzt natürlich voraus, dass Autoren bei der Contenterstellung auch Metadaten zu ihren Contentelementen hinzufügen, damit bei der Suche die relevanten Informationen gefunden werden. Um die Suche auch über mehrere Systeme hinweg möglich zu machen, sollten die Metadaten in einem zentralen Repository indiziert und für die Suche optimiert gehalten werden.

5.5 Benachrichtigungen

Wenn wieder verwendeter Content geändert wird, sollten Autoren, die diesen Content eingebunden haben, darüber benachrichtigt werden. Diese Benachrichtigung (Notifikation) wird durch einen Workflow-Prozess unterstützt und erfolgt in der Regel per Email. Eine Workflow-Applikation kann so definiert werden, dass jeder beliebige Autor informiert wird, wenn sich bestimmte Contentteile ändern. Der Notifikations-Prozess kann den Autor fragen, ob er die Änderungen akzeptiert oder nicht. Werden die Änderungen akzeptiert, so wird der geänderte Content automatisch nachgezogen. Bei Ablehnung kann der Autor die Verbindung zum wieder verwendeten Content lösen und dementsprechend seinen eigenen Content erstellen. Er muss nur noch Metadaten hinzufügen, und schon gibt es einen neuen Contentteil, der wiederum wieder verwendet werden kann.
In manchen Situationen werden keine Benachrichtigungen gewünscht sein, und Änderungen automatisch eingefügt werden müssen. Zum Beispiel bei Änderungen beim Firmenlogo oder bei Copyright- und Trademark-Informationen.

5.6 Erweitertes Rollen- und Rechtekonzept

Wenn Content in einem verteilten Content Management System wieder verwendet werden soll, muss man sich auch Gedanken über Zugriffsrechte machen. Nicht jeder Benutzer darf beliebigen Content wieder verwenden, sondern der Zugriff auf wieder verwendbaren Content sollte konfigurierbar sein. Da jedes einzelne System in den meisten Fällen die Benutzer der anderen Systeme nicht kennt, und jedes System sein eigenes Sicherheitsmodell beinhaltet, sollte man diese verschiedenen Sicherheitsmodelle vereinheitlichen, indem man eine einzige Schnittstelle den Systemen vorschaltet. Dieses Modell der Sicherheitsschnittstelle wird als „Single Sign On“ bezeichnet, und ist von einer Benutzer-Session und einem verbundenen Webserver, der diese

Session beinhaltet, abhängig. Für das Abrufen und Bearbeiten von wieder verwendbarem Content, wo oft Daten unabhängig einer Session verändert werden müssen, ist dieser Ansatz nicht ausreichend. Deswegen gibt es Entwicklungen für Sicherheitsstandards, wo die Daten mit Sicherheitsinformationen angereichert werden und mit den Daten verbunden sind. Die Security Assertion Markup Language (SAML) von der Firma OASIS[1] ermöglicht Sicherheitsinformationen als XML-Instanz abzulegen, die mit den Daten „mitwandert". Aussagen wie, wer (Benutzerrolle) hat Zugriff auf die Daten und wann (Prozessschritt) dürfen Daten verändert bzw. wieder verwendet werden, werden durch das System und die Daten verwaltet. *[Wald04]*

5.7 Grafische Eingabetools

Autoren müssen bei der Erstellung von strukturiertem Content so gut wie möglich durch bestimmte Editoren oder Tools unterstützt werden, damit keine Inkonsistenzen und Ungenauigkeiten auftreten. Jedoch sollte der Umgang mit diesen Tools auch intuitiv und leicht und schnell zu erlernen sein. Im Folgenden werden ein paar Möglichkeiten näher erläutert.

5.7.1 XML-Editoren

XML-Editoren wie zum Beispiel Epic oder XMetaL unterstützen Autoren bei der Erstellung von strukturiertem Content, indem sie das erstellte XML-Dokument anhand einer definierten DTD überprüfen, ob der erstellte Content gültig ist oder nicht. Ein Dokument ist dann gültig, wenn alle Regeln einer DTD erfüllt sind. Abbildung 5.5 zeigt ein Beispiel einer Fehlermeldung, wenn eine Regel verletzt wurde.

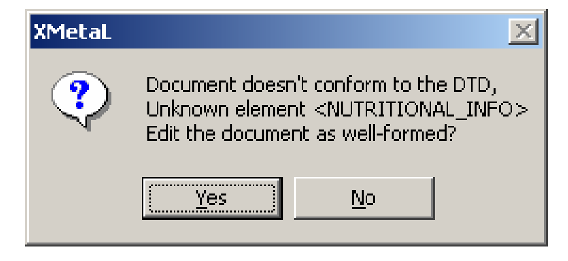


Abbildung 5.5 XMetaL-Meldung
Quelle: [Hack02]

[1] http://www.oasis-open.org

Weiters bieten solche Editoren mehrere Sichten auf XML-Dateien, wie reine XML-Ansicht, Tags-On Modus oder eine WYSIWYG-Ansicht, die stark Textverarbeitungsumgebungen gleicht. Auch eigene Stylesheets für das Eingeben und Ändern von Content können definiert werden. *[Hack02]*

5.7.2 Microsoft Word 2003

Microsoft hat in der neuen Office-Version 2003 einige Möglichkeiten eingebaut, um strukturierte Inhalte mithilfe von XML-Unterstützung zu erstellen. Word mit XML ist ein komplett unterschiedlicher Editor im Gegensatz zu den oben erwähnten XML-Editoren, denn es können sowohl alle Funktionen von Word, die uns schon aus früheren Versionen vertraut sind, uneingeschränkt genutzt werden, als auch ein paar zusätzliche Funktionen im Zusammenhang mit strukturiertem Inhalt und XML. Grundsätzlich kann Word 2003 nur eine limitierte Anzahl an XML-Strukturen behandeln, jedoch können eigene Applikationen für jeden Dokumenttyp im Unternehmen entwickelt werden. DMSI, Inc.[1] entwickelte so eine Applikation, mit dessen Hilfe man Dokumente erstellen und editieren kann, die auf den Regeln der DocBook DTD basieren.

Abbildung 5.6 zeigt Microsoft Word SmartDoc mit XML-Tags, wobei diese Ansicht eher einem klassischen XML-Editor entspricht. *[Wald04]*

[1] http://www.dmsi-world.com

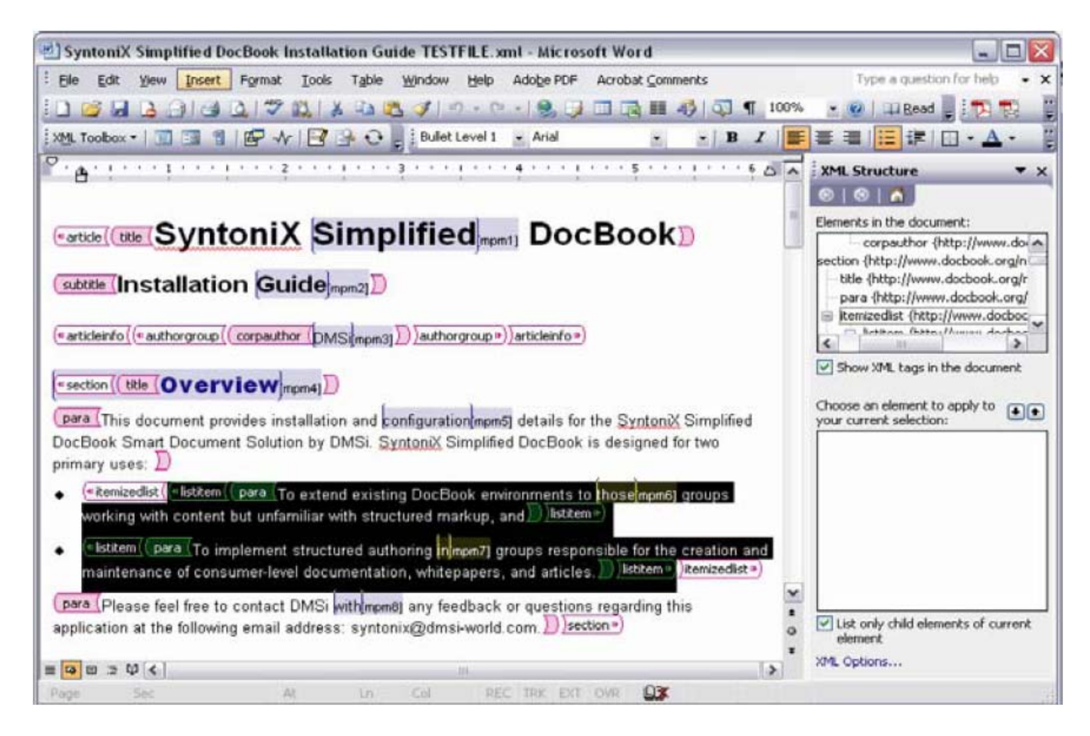


Abbildung 5.6 MS Word SmartDoc mit XML-Tags
Quelle: [Wald04]

Im Gegensatz dazu zeigt Abbildung 5.7 die Ansicht ohne XML-Tags, die nur die normale Stylesheet-Formatierung zeigt.

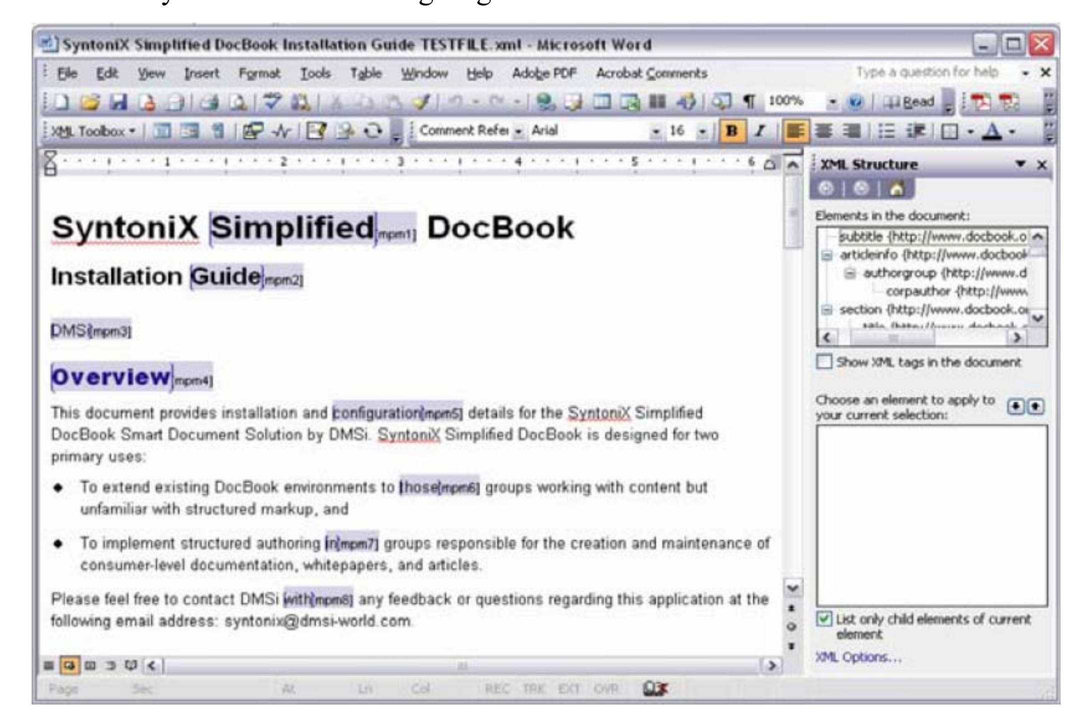


Abbildung 5.7 MS Word SmartDoc ohne XML-Tags
Quelle: [Wald04]

5.7.3 Web-Formulare

Eine weitere Möglichkeit, um strukturierten Content zu erstellen, kann über Web-Formulare realisiert werden. Ein Formular kann automatisch aus einer DTD generiert werden und den Autoren über eine Web-Oberfläche zur Verfügung gestellt werden. Dadurch ist die komplette XML-Syntax verborgen, und Autoren können sich beim Editieren von Inhalten auf das Wesentliche konzentrieren und werden nicht mit unnötigen Strukturinformationen belastet. Abbildung 5.8 zeigt wie ein Formular aussehen könnte, das aufgrund einer DTD generiert wurde. Die -/+ Buttons werden verwendet, um Steuerelemente am Formular und als Ergebnis auch Elemente im XML-Dokument hinzuzufügen oder wegzunehmen. *[Weit01]*

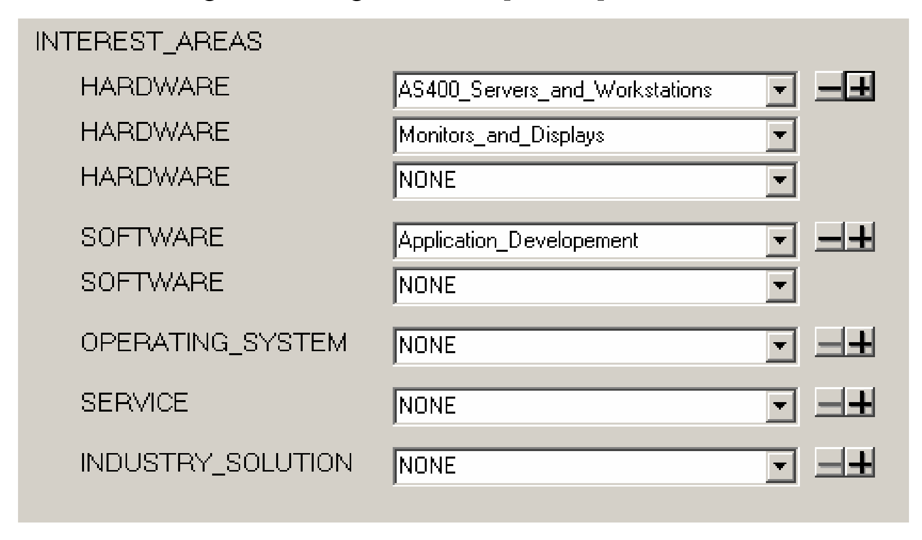


Abbildung 5.8 Generiertes Web-Formular
Quelle: [Weit01]

Dieser Ansatz nimmt die Funktionalität der Contenterfassung von den Client-Systemen und schiebt sie auf den Server. Somit müssen weder Installationen auf den Clients durchgeführt werden noch teure Software-Lizenzen eingekauft werden, und der gesamte Contentbearbeitungsprozess kann plattformunabhängig durchgeführt werden.

5.8 Interaktion der Systeme

Um Content in verteilten Content Management Systemen wieder verwenden zu können, müssen alle Systeme miteinander kommunizieren und Daten untereinander aus-

tauschen können. Es gibt Möglichkeiten, wie die Interaktion dieser Systeme über bestimmte Schnittstellen und Datenformate realisiert werden kann. Jedoch wird man bei jeder der Varianten über Anpassungen und Eigenimplementierungen nicht hinweg kommen, auch wenn alle Systeme im verteilten Content Management System gleich sind. Denn, um Content Wiederverwendung realisieren zu können, müssen die oben erwähnten Eigenschaften wie eine einheitliche Contentstruktur, erweiterte Metadaten, Abhängigkeiten, Suchfunktionen, Benachrichtigungen, erweiterte Zugriffsrechte und grafische Eingabetools zu jedem CMS hinzugefügt (dazuprogrammiert) werden. Je nach eingesetzter Technologie und verwendeter Datenaustauschformate bzw. nach den eingebauten Möglichkeiten und Funktionen des verwendeten Content Management Systems sind mehr oder weniger Anpassungen notwendig.

5.8.1 Import- und Exportschnittstelle

Über die Import- und Exportschnittstelle des Content Management Systems können Daten zwischen den Systemen ausgetauscht werden. Hierbei ist jedoch darauf zu achten, dass diese Schnittstellen unterschiedliche Formate behandeln können, sofern unterschiedliche Content Management Systeme eingesetzt werden. Hier werden sicherlich auch Anpassungen an den Schnittstellen erforderlich sein.

5.8.2 XML-Schnittstelle

Fast jedes Content Management System bietet eine XML-Schnittstelle oder XML-Unterstützung an. Diese Schnittstellen können optimal dazu verwendet werden, um Daten zwischen den Systemen auszutauschen, da XML, wie unter 4.4 beschrieben, beste Voraussetzungen für einheitliche Contentstrukturen, Metadaten, Formatierung, etc. bietet und zusätzlich noch plattformunabhängig ist.
Die Internet Technology Group von IBM hat auch schon einen Prototyp eines Content Management Systems entwickelt, der auf XML und XSL basiert und Content Wiederverwendung unterstützt. Nähere Informationen findet man unter *[Weit01]* und *[Weit02]*.

5.8.3 Webservices

In den letzten Jahren haben sich Webservices immer mehr durchgesetzt und der Begriff der „Service-oriented Architecture“ (SOA) ist in aller Munde. Webservices sind – einfach gesagt – Applikationen, die mithilfe von Informationsstandards die Interaktion von Systemen ermöglichen, wobei die Systeme selbst nicht bekannt sein müssen. Webservices basieren auf den offenen Standards SOAP (Simple Object Access Protocol), UDDI (Universal Description, Discovery and Integration) und WSDL (Web Services Description Language). Für die Content Wiederverwendung sind Webservices ein alternativer Ansatz, um Daten zwischen heterogenen Systemen auszutauschen. Jedes System stellt Funktionen für die Content Wiederverwendung als Webservices zur Verfügung, die dann von den anderen Systemen direkt oder über ein Zentralsystem aufgerufen werden können. Webservices sind an keine bestimmten Technologien gebunden und können somit über mehrere Systeme und Plattformen hinweg eingesetzt werden.

Kapitel 6

Bestehende CMS mit Funktionen für Content Wiederverwendung

Da es eine unüberschaubare Menge an derzeit am Markt erhältlichen Content Management Systemen gibt, wurden ein paar der meist verwendeten Systeme für eine Evaluierung im Hinblick auf allgemeine Informationen, eingesetzte Softwarekomponenten, Aspekte der Systemarchitektur und Möglichkeiten von Content Wiederverwendung herangezogen. Untersucht wurden sowohl Open-Source Produkte (OpenCms, Typo3) als auch hochwertige Enterprise Content Management Systeme (Interwoven TeamSite, Microsoft Content Management Server). Die Evaluierung beruht zur Gänze auf Produktdokumentationen und Onlinehilfen. Beim Microsoft Content Management Server wird auch ein selbst erstelltes Beispiel vorgestellt.

6.1 OpenCms

Das Content Management System „OpenCms“ wird von der Firma Alkacon Software kostenlos, unter der GNU Lesser Public License, auf der Webseite `www.opencms.org` angeboten. Die ersten Versionen des seit 1994 in Entwicklung befindlichen CMS erstellte und veröffentlichte Alexander Kandzior. Der erste Open Source Release Termin von OpenCms war im März 2000. Seit dem Jahr 2002 wird das CMS von Alkacon Software, einer von Kandzior gegründeten Firma, weiterentwickelt und publiziert. Die aktuelle Version von OpenCms ist 6.0.3. *[Open]*

6.1.1 Systemarchitektur

OpenCms ist eine Client-Server-Applikation und basiert auf folgenden Softwarekomponenten *[Cmsm]*:

Betriebssysteme:	plattformunabhängig
Programmiersprachen:	Java 1.3+, XML
Webserver:	Apache, IIS
Applikationsserver:	Tomcat, JBoss, JRun, ...
Datenbanken:	MySQL, PostGreSQL, Oracle, MSSQL
Lizenzkosten:	keine

Das Herzstück von OpenCms ist der so genannte OpenCms Arbeitsplatz, der aus drei Teilen besteht *[Bitt04]*:

- Explorer-Ansicht
  Dient dem Erstellen und Bearbeiten von Ordnern, Templates, HTML-Seiten, etc.
- Administrator-Ansicht
  Dient der Administration von Projekten, Benutzern, Gruppen, Datenbank, Modulen, Ressourcen, etc.
- Workflow-Ansicht
  Dient dem Erstellen und Bearbeiten von Aufgaben und Benachrichtigungen.

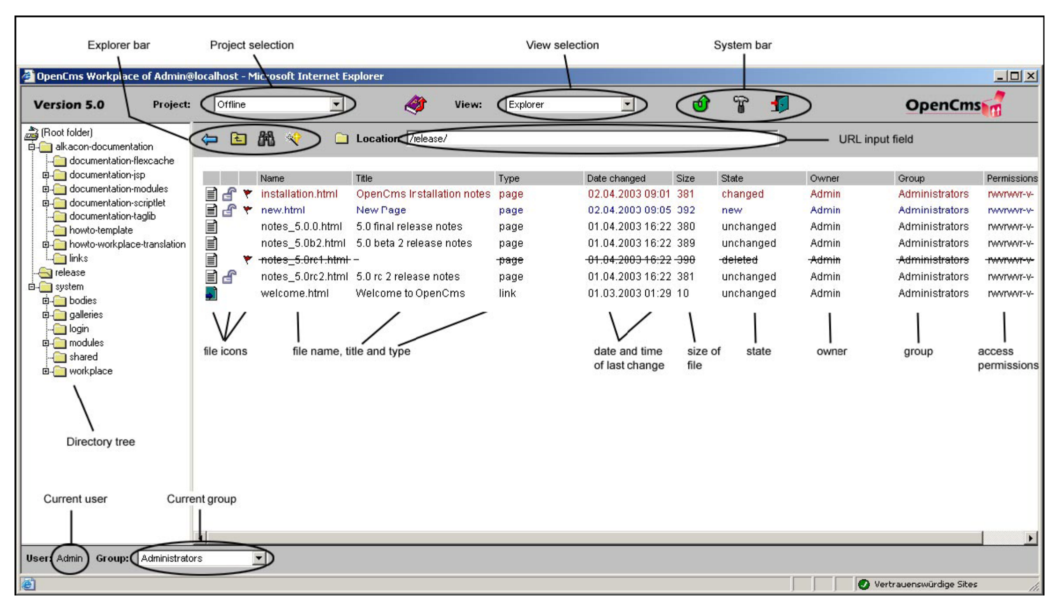


Abbildung 6.1 OpenCms Arbeitsplatz – Explorer-Ansicht
Quelle: [Alka03a]

OpenCms bietet eine projektbasierte Arbeitsweise, wobei Projekte in Online- und Offline-Projekte unterteilt sind. Änderungen werden nur an Offline-Projekten durchgeführt und vom Projektmanager veröffentlicht. Alte Versionen aller Dateien werden in der History gespeichert und sind jederzeit wieder herstellbar.

OpenCms verwendet einen flexiblen Templatemechanismus, um Seiten mithilfe von Elementen zusammenzubauen. Jede Seite ist aus einem Mastertemplate, Frametemplate und Contenttemplate zusammengesetzt. Das Mastertemplate enthält Metadaten über die Templates, das Frametemplate beschreibt das Design der Seite und das Contenttemplate beschreibt das Layout des Contents der Seite und kann aus mehreren Body-Elementen bestehen. Ein Body-Element enthält den tatsächlichen Content, der durch einen WYSIWYG-Editor im Browser bearbeitet werden kann. Templates können proprietär mithilfe von XML erstellt werden, jedoch wird seit Version 5 empfohlen, Templates mittels JSP zu entwickeln. Elemente, die in unterschiedlichen Templatesprachen erstellt wurden, können auch miteinander verbunden werden.

OpenCms beinhaltet einen Modulmechanismus, der einen leichten und bequemen Weg bietet, um Templates und andere Funktionalitäten bündeln und verteilen zu können. Ein Modul ist gewöhnlich eine zusammenhängende Menge von HTML-Seiten, Templates, Stylesheets, Bildern und Java-Klassen, die eine ganz bestimmte Funktionalität bereitstellen. *[Alka03a]*

6.1.2 Content Wiederverwendung

Content Wiederverwendung kann über so genannte Content Definitionen realisiert werden. Eine Content Definition ist eine Klasse, die eine Contentstruktur definiert und sämtliche Methoden für Datenzugriff und Manipulation der Daten, mitunter auch Datenbankzugriffe, etc. beinhaltet. Wenn Methoden zur Contentanzeige von verschiedenen Templates aufgerufen werden, wird derselbe Content auf jeder Seite angezeigt. Weitere Informationen über „Content Definitionen“ findet man in der Dokumentation von OpenCms unter *[Alka03a]*.

Ab OpenCms 6.0 kann Content Wiederverwendung auch über die Eigenschaft XML Content realisiert werden. Durch XML Content kann man einfach strukturierten Content wie Nachrichten, Adressdaten, Listenansichten, etc. erstellen. XML Content

wird im virtuellen Dateisystem von OpenCms abgelegt und kann in einem automatisch generierten formularbasierten Editor bearbeitet werden. Über eine XSD (XML Schema Definition) kann die Contentstruktur festgelegt und Mehrsprachigkeit, Validierungen, Defaultwerte, etc. definiert werden. Wird dann derselbe XML Content in verschiedene JSP-Templates eingebunden, wird dieser Content automatisch wieder verwendet. Weitere Informationen findet man in der interaktiven Online Dokumentation „OpenCms XML content documentation", die mit OpenCms mitgeliefert wird.

Diese beiden Arten der Content Wiederverwendung sind nur innerhalb derselben Installation von OpenCms anwendbar und können weder für gesamte Navigationsbäume noch für gesamte Seiten verwendet werden. Es werden lediglich gewisse Contentelemente, die auf bestimmten Contenttypen basieren, wieder verwendet.

Will man Content Wiederverwendung in größerem Ausmaß betreiben, kommt man nicht darüber hinweg, eigene Module zu implementieren. Über die Modulimportschnittstelle werden diese Module ins System integriert und können von Administratoren jederzeit konfiguriert werden. Mithilfe dieser selbsterstellten Module kann man auch Import- und Exportschnittstellen des Systems bedienen, sowie ganze Seiten bzw. sogar Navigationsbäume wieder verwenden. Weiters können auch Änderungsnotifikationen implementiert werden. Diese Art der Content Wiederverwendung setzt jedoch gute Programmierkenntnisse voraus und bedeutet sehr großen Implementierungsaufwand.

6.2 Typo3

Typo3 ist ein Open Source Content Management System und wird seit 1997 von dem Dänen Kasper Skårhøj mit der Hilfe und Anregung von Usern entwickelt. Typo3 wird unter der General Public License, auf der Webseite `www.typo3.org` angeboten. Die aktuelle Version von Typo3 ist 3.8.1. *[Typo]*

6.2.1 Systemarchitektur

Typo3 basiert auf folgenden Softwarekomponenten *[Cmsm]*:

Betriebssysteme:	plattformunabhängig
Programmiersprachen:	PHP
Webserver:	Apache, IIS

Applikationsserver:	PHP 4.1.2+
Datenbanken:	MySQL
Lizenzkosten:	keine

Typo3 ist in ein Frontend und ein Backend eingeteilt, wobei beide Teile im Browser dargestellt werden. Im Backend werden administrative und redaktionelle Arbeiten wie zum Beispiel neue Seiten oder Templates anlegen, Benutzer verwalten, Erweiterungen (Module) einspielen durchgeführt. Die Backend-Arbeitsoberfläche von Typo3 ist in drei Hauptbereiche unterteilt *[Meye04]*:

- Modulleiste
  In der Modulleiste kann man das gewünschte Arbeitsmodul auswählen. z.B.: Web, Datei, Benutzer, Hilfe.
- Navigationsansicht
  In der Navigationsansicht wird die komplette Seitenstruktur baumartig visualisiert.
- Detailansicht
  Die Detailansicht (auch Datenansicht) stellt die eigentliche Arbeitsfläche für den Nutzer dar. In diesem Bereich werden die Details zu den in der Navigationsansicht ausgewählten Elementen angezeigt.

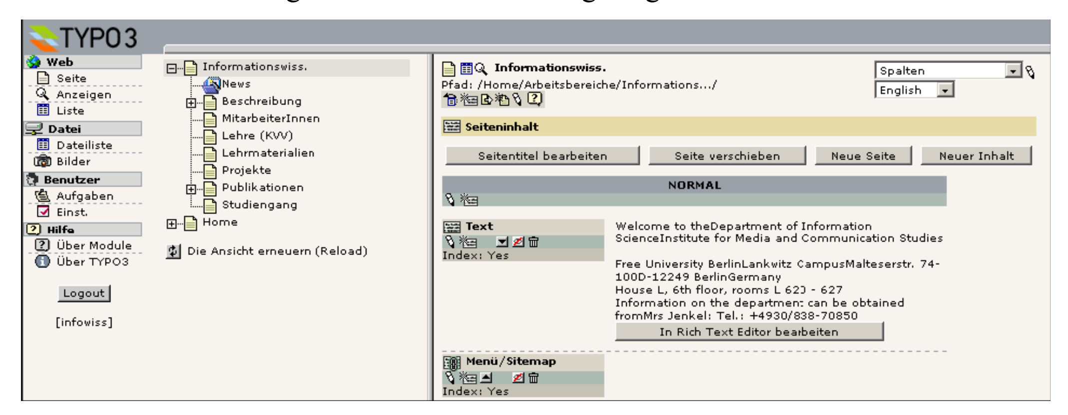


Abbildung 6.2 Typo3 Backend
Quelle: [Müll05]

Das Frontend dient nur zur Bearbeitung der Contentelemente der Seiten, und eignet sich nur für kleine Anpassungen und Änderungen wie zum Beispiel Schreibfehler.
Als Templates dienen in Typo3 HTML-Dateien, die im Code mit Platzhaltern („Subparts") versehen sind, die durch dynamischen Inhalt ersetzt werden. Zum Bearbeiten

von Templates kommt „TypoScript“, eine deklarative systemeigene Skriptsprache zum Einsatz, die der publishing engine angibt, wie welcher Datenbankinhalt angezeigt werden soll. Templates können in der Hierarchie vererbt und in Unterseiten erweitert werden oder können wiederum aus Templates bestehen.

Typo3 ist seit der Version 3.5.x vollständig modular aufgebaut. Eigene Modulentwicklungen lassen sich nun komfortabel installieren oder können in der so genannten „Extension Repository“ zentral abgelegt werden, und so der Öffentlichkeit zur eigenen Verwendung zur Verfügung gestellt werden. *[Müll05]*

6.2.2 Content Wiederverwendung

Content Wiederverwendung wird standardmäßig wie folgt unterstützt: Wenn man auf einer Seite einen neuen Contentbereich anlegen will, muss man den Typ des Contents aus einer Liste auswählen. Wählt man den Typ „Insert records“ aus, kann man diesen Contentbereich auf einen anderen Contentbereich des Systems verweisen lassen. Wird der Content auf der „Quellseite“ geändert, sind die Änderungen natürlich auch auf der „Zielseite“ sichtbar. Im Backend findet man sogar auf der Verweisseite einen Link auf den Content der verwiesenen Seite, um diesen auch aus dieser Seite heraus editieren zu können. *[Typo]*

Content Wiederverwendung ist auch über das Erweiterungs-Modul „Templavoila“ möglich. Templavoila dient zur einfachen Erstellung von Templates und eigenen Contenttypen und wird als Standard Templateeditor der Zukunft gesehen. Eine Eigenschaft von Templavoila ist, dass man zu jedem Contentelement eine Referenz auf ein anderes Contentelement im System anlegen kann. Durch diese Referenz ist Content Wiederverwendung realisiert, denn der Content wird nur einmal geschrieben und kann beliebig oft referenziert werden. *[Skar04]*

Diese beiden Arten der Content Wiederverwendung sind sehr ähnlich und können jeweils nur innerhalb desselben CMS eingesetzt werden. Auch Notifikationen an den Contentverfasser sind in diesen Versionen nicht möglich.

Wenn man speziellere Funktionen, wie Import- bzw. Exportfunktionen zu anderen Content Management Systemen oder Notifikationen für Content Wiederverwendung haben will, muss man sich sein eigenes Modul entwickeln, und dieses dann über die Erweiterungsschnittstelle einbinden.

6.3 Interwoven TeamSite

Die Firma Interwoven wurde 1995 gegründet, um Software für Website-Produktion auf den Markt zu bringen. Mit der Entwicklung von TeamSite entstand ein Produkt, um für große Unternehmen mit sehr riesigen, komplexen Websites (mehr als 1000 Seiten und mehrere Dutzend Benutzer) den Content-Erstellungsprozess zu unterstützen. Interwoven TeamSite ist in der aktuellen Version 6.5 erhältlich. *[Inte]*

6.3.1 Systemarchitektur

TeamSite basiert auf folgenden Softwarekomponenten *[Cmsm]*:

Betriebssysteme:	Linux, Solaris, Windows
Programmiersprachen:	Java, JSP, Perl, CGI, C++, JavaScript, HTML
Webserver:	Apache, IIS
Applikationsserver:	integriert
Datenbanken:	integriert
Lizenzkosten:	Standard-Edition: ca. 49 000 $/Jahr
	Enterprise-Edition: ca. 159 000 $/Jahr

Das Frontend besteht aus folgenden browserbasierten Bestandteilen *[Inte05a]*:

- ContentCenter Standard
  Ein Interface für gewöhnliche Benutzer und technisch nicht versierte Manager, um mithilfe von Assistenten und einfachen Möglichkeiten Content zu erstellen und zu veröffentlichen.
- ContentCenter Professional
  Ein erweitertes Interface für technische Benutzer, Projektmanager und Administratoren. Technische Benutzer können mit wenigen Klicks die meisten Content Management System Funktionalitäten ausführen, Projektmanager können auf schnelle und einfache Weise den Content-Veröffentlichungsprozess überwachen und Administratoren können Konfigurationseinstellungen ändern und Prozesse überwachen.
- FormsPublisher
  Der FormsPublisher ist ein formularbasiertes XML Interface mit einem

WYSIWYG-Editor, ad-hoc-Vorschau und diversen Anpassungsmöglichkeiten auf Formularbasis.

Abbildung 6.3 TeamSite ContentCenter Standard
Quelle: [Inte05a]

Die Konfigurations- und Anpassungsmöglichkeiten von TeamSite scheinen geradezu unendlich, sogar das ContentCenter-Interface ist den persönlichen Wünschen entsprechend konfigurierbar. TeamSite unterstützt Erweiterungen sowohl auf J2EE- als auch auf .NET-Plattformen und ist zu allen Komponenten der Enterprise Content Management Plattform von Interwoven voll kompatibel. So gibt es zum Beispiel eigene Komponenten für Dokumentenmanagement, Integration, Verteilung und Transformation von Content, uva. Weiters unterstützt die Service-orientierte Architektur (SOA) von TeamSite eine einfache und standardisierte Integration anderer

Business-Applikationen wie CRM (Customer Relationship Management) oder HR (Human Resources). *[Inte05a]*

Teamsite kann mehrere Websites auf demselben System verwalten. Die verwalteten Websites sind in so genannte „Branches“ unterteilt und beinhalten „Workareas“ und eine Staging-Umgebung. Workareas sind virtuelle Kopien eines Zweiges, auf die verschiedene Teams zugreifen können und Content abhängig von einer Contentbasis verändern können. Dabei ist jedoch – ähnlich einem Configurationmanagement-Tool – immer sicher gestellt, dass Content nicht überschrieben oder von zwei Teams gleichzeitig bearbeitet werden kann. Nach Content-Anpassungen stellen Autoren ihre Änderungen auf der Staging-Umgebung zur Verfügung, wo sie mit den Workareas anderer Teams zusammengefügt werden. Die Staging-Umgebung ermöglicht es, Fehler zu finden und zu beheben bevor die endgültige Version publiziert wird. *[Trut04]*

6.3.2 Content Wiederverwendung

Content Wiederverwendung kann über die Zusatzkomponente TeamSite XML Server realisiert werden. Autoren können XML Autorensysteme verwenden, um XML Komponenten zu erstellen, zu editieren, zu finden und wieder zu verwenden. Hierarchische Beziehungen zwischen den Komponenten bleiben erhalten, da alle wieder verwendeten Komponenten automatisch geändert werden, wenn die Basiskomponente verändert wird. Diese Art der Wiederverwendung kann im gesamten System, und somit auch über mehrere Websites hinweg eingesetzt werden. *[Inte05a]*

Content Wiederverwendung in Verbindung mit anderen Systemen kann über eine andere Zusatzkomponente realisiert werden. Mit der Komponente „Content Transformation Services“ wird Content als wieder verwendbares XML abgespeichert und über XSL in andere Formate wie HTML oder PDF umgewandelt. Administratoren können über eine mächtige grafische Oberfläche Content aus bestehenden Dokumenten extrahieren oder aber Content, der bestimmten XML Schema Anforderungen entsprechen soll, als XML abspeichern. Dieser XML Content kann dann auch Schnittstellen anderer Content Management Systeme bedienen. *[Inte05b]*

6.4 Microsoft Content Management Server

Microsoft hat 2001 das Content Management System der kanadischen Firma nCompass aufgekauft und vertreibt es nun unter eigenem Namen. Die aktuelle Version ist der Microsoft Content Management Server 2002 SP2.

6.4.1 Systemarchitektur

Der Microsoft Content Management Server (MCMS) basiert auf folgenden Softwarekomponenten *[Cmsm]*:

Betriebssysteme:	Windows 2000, 2003
Programmiersprachen:	.NET 1.0+
Webserver:	IIS
Applikationsserver:	IIS, ASP.NET
Datenbanken:	MSSQL
Lizenzkosten:	ca. 40 000 $/Jahr

Beim MCMS können Seiten nur im Frontend, d.h. im MS Internet Explorer neu angelegt bzw. editiert werden. Das Backend setzt sich aus drei Applikationen zusammen *[Engl04]*:

- Database Configuration Application (DCA)
  Die DCA dient der Administration der Datenbank.
- Server Configuration Application (SCA)
  Die SCA dient der Konfiguration des Servers wie Webserverzugang, Gastzugang, Adminaccounts
- Site Manager
  Im Site Manager werden Ressourcen, Templates und Benutzer verwaltet und auch die Navigationsstruktur wird festgelegt.



Abbildung 6.4 MCMS Frontend

Der Content Management Server integriert sich automatisch in das Visual Studio .NET, der Standard-Entwicklungsumgebung von Microsoft für .NET-Entwicklung. Templates können ausschließlich im MS Visual Studio erstellt werden und können beim Anlegen neuer Seiten im Frontend ausgewählt werden. Für jedes Template werden so genannte Content-Placeholder definiert, die den tatsächlichen Content beinhalten und im Frontend editiert werden können. Es gibt drei Standardplaceholder, jedoch können beliebige Arten von Placeholder definiert bzw. programmiert werden.

Erweiterungen können beliebig dazuprogrammiert werden bzw. können im Internet heruntergeladen werden. Weiters lässt sich der MCMS mit anderen Microsoft-Produkten kombinieren. So kann zum Beispiel der Commerce Server, der Webshop-funktionalitäten bereitstellt oder der Sharepoint Portal Server, der Funktionen für Dokumentenmanagement bietet, mit dem Content Management Server verbunden werden. *[Engl04]*

6.4.2 Content Wiederverwendung

Content Wiederverwendung ist realisiert über so genannte „Connected Pages". Connected Pages sind zwei oder mehr Seiten, die eine gewisse Menge an gleichem Content beinhalten. Wenn Content von einer dieser „verbundenen" Seiten verändert wird, werden automatisch alle verbundenen Seiten geändert. Miteinander verbundene Seiten können entweder dasselbe Template oder so genannte „Connected Templates" verwenden. Connected Templates haben den Vorteil, dass verschiedene Layouts verwendet werden können, der Content anders dargestellt werden kann oder nur ein Teil des Contents wieder verwendet werden kann. Diese Art der Wiederverwendung kann nur innerhalb desselben Content Management Systems verwendet werden, jedoch aufgrund der Multisite-Fähigkeit des MCMS können Connected Pages auch auf verschiedenen Sites verwendet werden. Email-Notifikationen sind standardmäßig nicht vorgesehen, jedoch können diese dazuprogrammiert werden.
Um Content auch in oder von anderen CMS wieder zu verwenden, kann Content mithilfe von XML-Placeholder auch als XML zur Verfügung gestellt werden oder es können eigene Contenttypen definiert werden.
Weiters bietet das .NET Framework und die API des MCMS genügend Funktionalitäten, um die jeweils gewünschte Art von Content Wiederverwendung programmatisch realisieren zu können. *[Engl04]*

6.4.3 Beispiel

Das Beispiel zeigt, wie man beim Microsoft Content Management Server anhand von „Connected Pages" Content wieder verwenden kann.

Zuerst muss man eine neue Seite mit Content, der wieder verwendet werden soll, anlegen. Abbildung 6.5 zeigt diese Seite, die auf einem Template basiert, das oben die Primärnavigation, darunter die Breadcrumb-Navigation und auf der linken Seite die Sekundärnavigation beinhaltet. Der Content in der Mitte der Seite besteht sowohl aus einer Grafik als auch aus formatiertem Text. Auf der rechten Seite befindet sich die so genannte CMS-Console, die die CMS-Funktionen wie Anlegen, Editieren und Veröffentlichen der Seite, Anzeigen und Ändern von Seiteneigenschaften und Assistenten für Seitenfreigaben und Ressourcen bereitstellt.

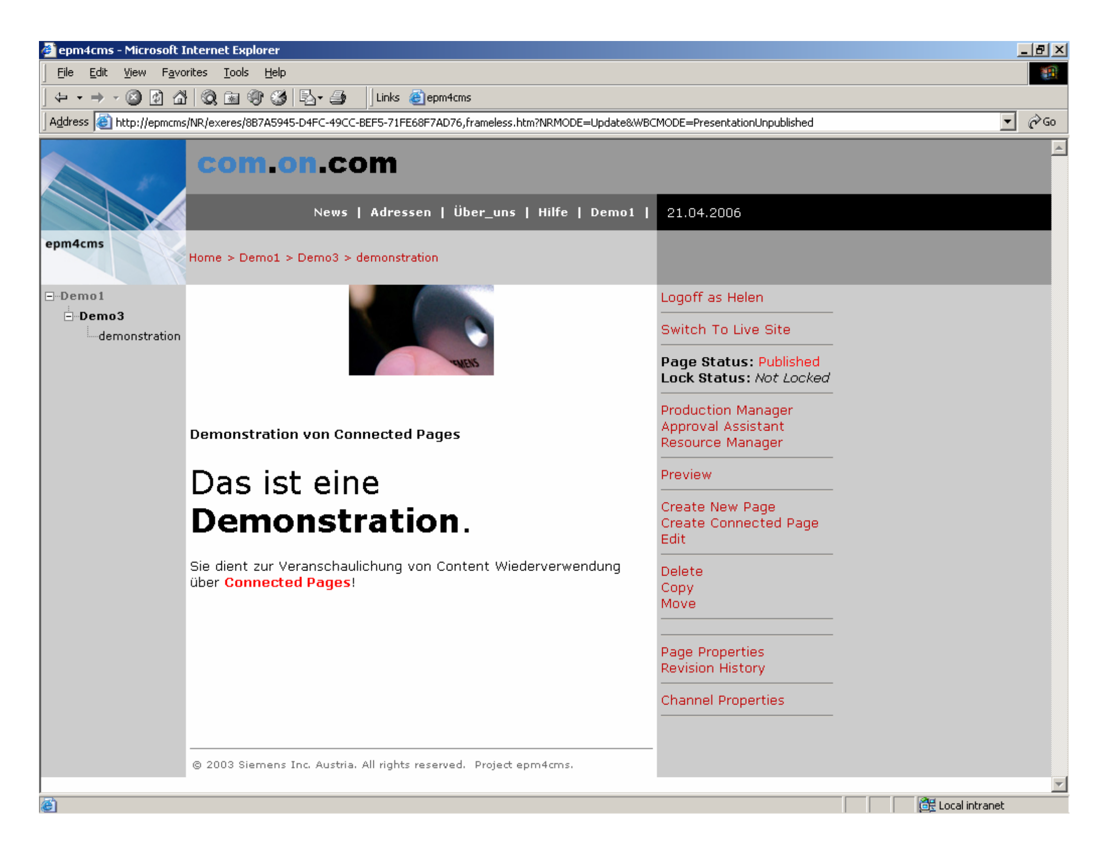


Abbildung 6.5 MCMS „Connected Page“

Nun muss man eine „Connected Page“ anlegen. Beim Anlegen der Seite muss man den Ort angeben, an dem die Seite abgelegt werden soll. Abbildung 6.6 zeigt dieses Auswahlfenster.

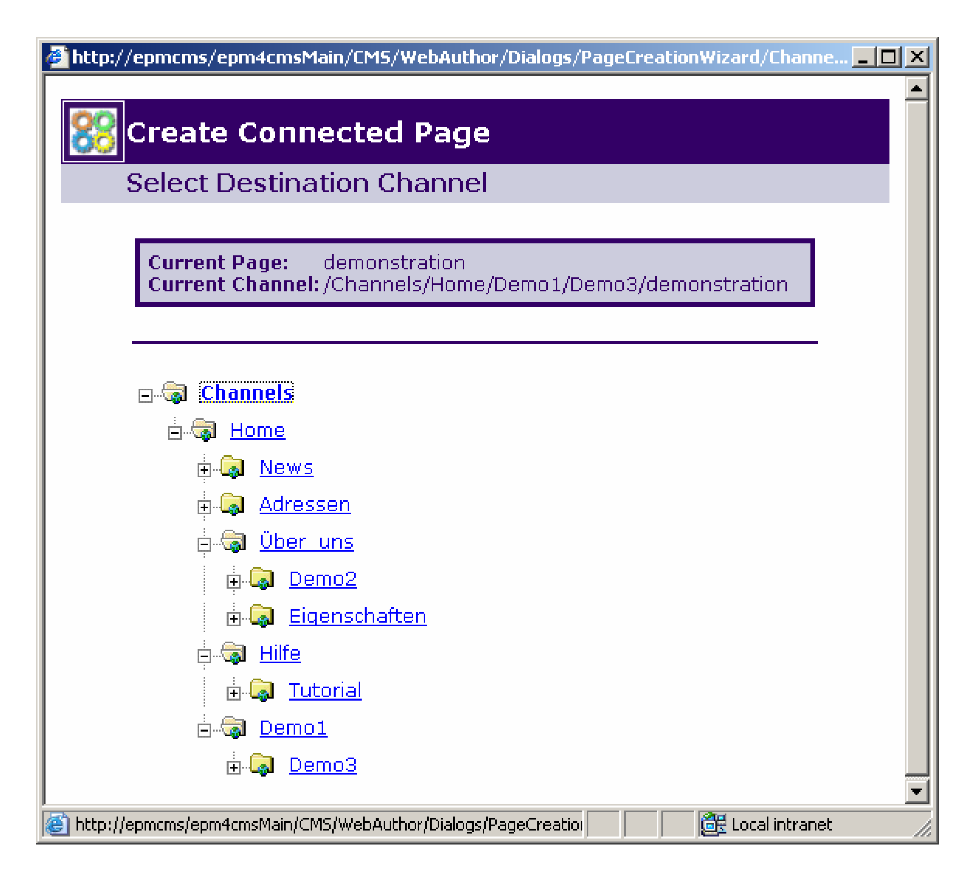


Abbildung 6.6 MCMS „Connected Page" – Zielort-Auswahl

Dann muss man das Template auswählen, das für diese Seite verwendet werden soll. Für verbundene Seiten können nur zwei Arten von Templates verwendet werden. Einerseits kann für alle Seiten dasselbe Template verwendet werden und andererseits können so genannte „Connected Templates" verwendet werden. Connected Templates haben den Vorteil, dass nicht der gesamte Content einer Seite wieder verwendet wird, und somit nicht eine Seite genau einer verbundenen Seite gleicht. So können Contentteile anders angeordnet werden, nur bestimmte Teile des Contents wieder verwendet werden oder Content anders formatiert werden. In diesem Beispiel wird nur dasselbe Template zum Auswählen bereitgestellt, da keine Connected Templates im System vorhanden sind. Abbildung 6.7 zeigt das Auswahlfenster für die Templates.

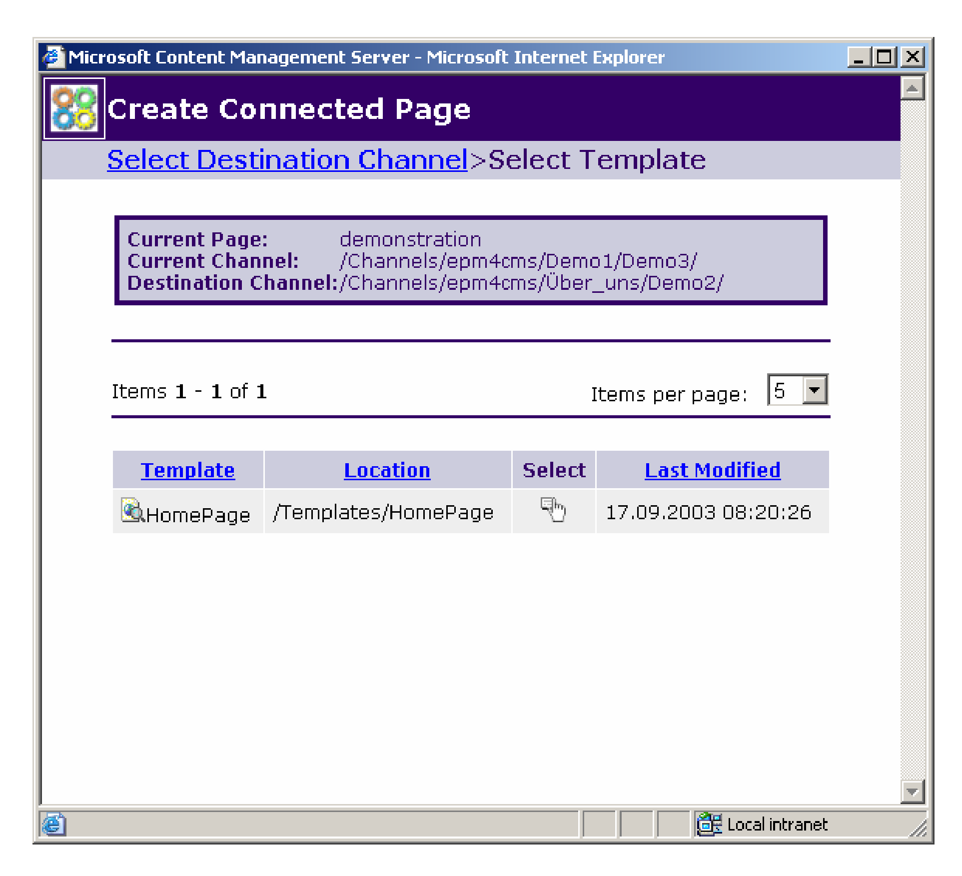


Abbildung 6.7 MCMS „Connected Page“ – Template-Auswahl

Beim Speichern muss man der Seite noch einen Displaynamen geben, der in der Navigation sichtbar ist. Der wirkliche Name, der zur internen Identifizierung der Seite dient, ist bei allen Connected Pages identisch, nur der Name, der angezeigt wird, ist unterschiedlich.

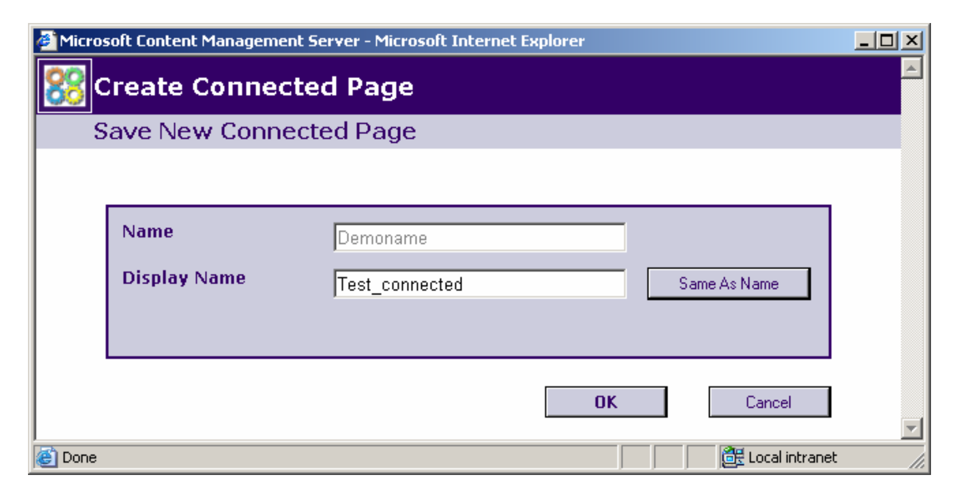


Abbildung 6.8 MCMS „Connected Page“ – Speichern

Weiters bietet das CMS von Microsoft eine Funktion, um alle Connected Pages zu einer Seite anzeigen zu können. Abbildung 6.9 zeigt dieses Fenster, wobei diese Seite nun schon drei verbundene Seiten hat. Die Darstellung beinhaltet den Displaynamen und den Ort der verbundenen Seite sowie einen Link zu dieser Seite.

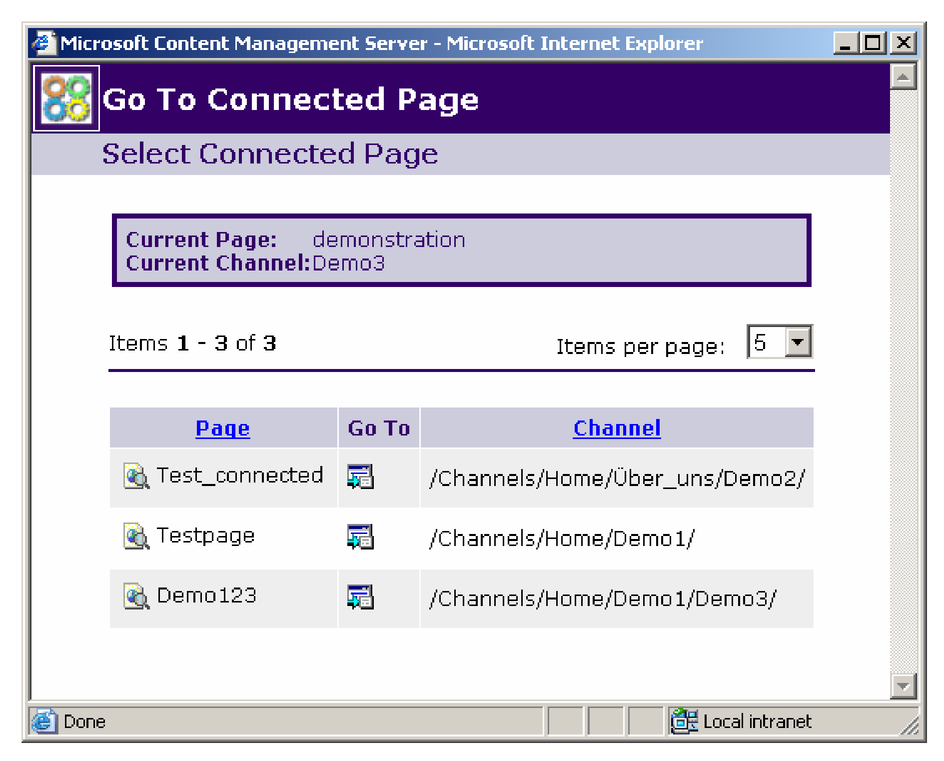


Abbildung 6.9 MCMS „Connected Page“ – Navigieren zu verbundenen Seiten

Wird nun Content auf irgendeiner Seite verändert, werden die Änderungen automatisch auf allen Connected Pages nachgezogen. Jedoch müssen diese Änderungen auf jeder dieser Seiten wiederum freigegeben werden.

Es gibt einige Regeln, die vom Content Management System im Zusammenhang mit Connected Pages eingehalten werden, und auf die man bei der Erstellung von Connected Pages achten sollte *[Engl04]*:

- Eine Connected Page kann nicht im selben Navigationszweig angelegt werden, wenn dasselbe Template verwendet wird.
- Alle Eigenschaften wie zum Beispiel wann eine Seite online sichtbar ist, der Displayname oder selbst definierte Eigenschaften von Connected Pages sind pro Seite eigens definierbar. Ausgenommen davon sind der wirkliche Name

und die Seitenbeschreibung. Diese Eigenschaften sind für alle verbundenen Seiten gleich.

- Ein Autor, der Connected Pages erstellt, muss auch für die Speicherorte der verbundenen Seiten Schreibrechte besitzen.
- Es herrscht keine Parent-Child-Beziehung (siehe 5.3.1) zwischen den Seiten, sondern alle Seiten sind einander gleichgestellt.

Connected Pages bieten eine rudimentäre Lösung, um Content wieder zu verwenden und werden vor allem für Mehrsprachigkeit eingesetzt. Wenn man mehr Funktionalitäten für Content Wiederverwendung nutzen möchte, müssen diese über die Programmierschnittstelle dazu implementiert werden.

6.5 Auswertung

Alle untersuchten Content Management Systeme bis auf Interwoven TeamSite bieten standardmäßig eine Art von Content Wiederverwendung an. In TeamSite kann man Content nur wieder verwenden, wenn man Zusatzkomponenten einbindet. Die Standardfunktionen für Wiederverwendung sind bei Typo3 und beim Microsoft Content Management Server am intuitivsten und sehr leicht anwendbar.
Jedes System stellt Möglichkeiten und Funktionen zur Verfügung, um Inhalte innerhalb des Systems wieder zu verwenden. Doch mithilfe von XML ist Content Wiederverwendung bei allen Systemen außer Typo3 auch über Systemgrenzen hinweg anwendbar. Typo3 bietet als einziges System keinen Content in XML-Form bzw. Wiederverwendung durch XML-Funktionen.
Es gibt kein untersuchtes System, das out-of-the-box ganze Navigationsbäume wieder verwenden kann oder Notifikationen für Contentverfasser bereitstellt. Jedoch kann jedes System über seine API durch mehr oder weniger großen Programmieraufwand erweitert werden. Systemerweiterungen sind bei den beiden Open-Source Produkten dank ihrer großen Community im Internet zur Genüge frei verfügbar. Auch Microsoft und Interwoven bieten Erweiterungen und Hilfestellungen im Internet zum Download an.
Abschließend ist zu sagen, dass jedes System standardmäßig auf mehr oder weniger einfache Art und Weise Content Wiederverwendung anbietet. Jedoch kann bei jedem

untersuchten System mithilfe von XML und spezifischen Erweiterungen jede beliebige Art der Content Wiederverwendung implementiert werden. Meiner Meinung nach bietet Interwoven TeamSite mit allen Zusatzkomponenten die besten Möglichkeiten, um Content wieder zu verwenden.

6.6 Persönliche Erfahrung

Durch meine Erfahrung mit dem CMS von Microsoft kann ich sagen, dass standardmäßig viele Anpassungen und Eigenentwicklungen notwendig sind, um ein CMS nach den gewünschten Anforderungen einsetzen zu können. Wenn man dieses Content Management System installiert hat, hat man nicht mehr als das Application Programing Interface (API) zur Verfügung. Man muss im Prinzip von Null weg alles implementieren, um das CMS nutzen zu können. Jedoch dank mehrerer Beispielseiten (sample sites), die zum Teil mit dem Produkt ausgeliefert werden oder im Internet frei herunterladbar sind, stehen sehr viele Funktionen schon vorgefertigt zur Verfügung. So kann man zum Beispiel Code für Primär- und Sekundärnavigation, Breadcrumb-Navigation oder spezifische CMS-Funktionalitäten wie Anlegen, Editieren und Veröffentlichen von Seiten, sofort verwenden. Weiters gibt es dank einer großen Internet-Community sehr viele Codefragmente für die verschiedensten Anwendungsgebiete, wie zum Beispiel für XML-Content oder Email-Notifikationen, frei zum Herunterladen. Für eine einfache Lösung müssen dann eigentlich nur mehr das Layout der Templates und die Anordnung der verschiedenen Funktionen auf der Seite programmiert werden. Will man eigene Datenbanken an das CMS anbinden oder spezielle Funktionen auf den Seiten bereitstellen, kommt man um Individualprogrammierungen ohnehin nicht herum. Das heißt, je komplexer und spezieller die Anforderungen einer Anwendung sind, desto weniger können vorgefertigte Elemente verwendet werden und der Aufwand für Anpassungen und Eigenentwicklungen steigt.

Kapitel 7

Zusammenfassung

Content Wiederverwendung in verteilten Content Management Systemen besitzt enormes Potential, um den Contenterstellungsprozess zu erleichtern und den Content eines Unternehmens konsistent und qualitativ hochwertig zu halten. Jedoch ist die Einführung von Content Wiederverwendung in einem Unternehmen auch ein sehr kritischer Prozess. Die Komplexität von Content Wiederverwendung hat natürlich auch Einfluss auf die Usability und Bedienbarkeit des CMS. Für Autoren ist es besonders schwierig von einem seitenbasierten Modell, wo jede Seite im CMS der veröffentlichten Seite gleicht, zu einem komponentenbasierten Modell, wo eine Seite aus mehreren Contentkomponenten, die in verschiedenen getrennten Strukturen verwaltet werden, zu wechseln. Ein weiterer wichtiger Aspekt sind Contentupdates. Wenn einmal Content Wiederverwendung eingesetzt wird, müssen Autoren sehr sorgfältig bei Änderungen des Contents sein. Ohne Sorgfalt und hinreichende Unterstützung des CMS, können Autoren leicht Fehler machen, die sehr viele Stellen betreffen. Auch ist es oft leichter, Content Wiederverwendung in einem zentralisierten System einzusetzen, wo die Autoren eng bei der Verwaltung der Beziehungen zwischen den Seiten zusammenarbeiten. Deswegen wird Content Wiederverwendung in einem dezentralen System, wo eine große Menge an Autoren voneinander unabhängig arbeitet, zu einer richtigen Herausforderung. In so einer Situation wird wahrscheinlich Content kaum wieder verwendet werden, da Autoren oft gar nicht wissen, dass wieder verwendbarerer Content existiert. Aus diesen Gründen muss sehr viel

Aufwand in das Design des CMS gesteckt werden, damit die Benutzer des Systems die Möglichkeiten von Content Wiederverwendung verstehen und nutzen können. Content kann nur dann effektiv wieder verwendet werden, wenn die Wiederverwendungsstrategie sorgsam implementiert und das Design sauber durchdacht wurde.

Content Wiederverwendung bietet zwar einigen Nutzen, jedoch wird sie nicht alle Probleme bezüglich Qualität, Genauigkeit und Konsistenz des Contents lösen. Bei der Auswahl und Implementierung des Content Management Systems sollte man Rücksicht auf diese Aspekte nehmen. Damit ist sichergestellt, dass die anderen Vorteile des CMS nicht verloren gehen, wenn man sich zu sehr auf Content Wiederverwendung konzentriert. Bevor man die Möglichkeiten von Content Wiederverwendung von Content Management Systemen betrachtet, sollte man zuerst die Abteilungen im Unternehmen identifizieren, wo Wiederverwendung überhaupt möglich und machbar ist. Nicht in jeder Abteilung ist Content Wiederverwendung einsetzbar, da es oft zu große Unterschiede in der Struktur und Beschaffenheit des Contents gibt. Das Einführen von Content Wiederverwendung in einem Unternehmen verlangt grundlegende Änderungen in den Businessprozessen, und verursacht somit sehr viel Zeitaufwand und Kosten. Deswegen sollte man eine Umstellung nicht auf einmal für das gesamte Unternehmen durchführen, sondern man sollte mit den Abteilungen beginnen, die am meisten durch Content Wiederverwendung profitieren, und sich dann sukzessive weiter durcharbeiten.

Content Wiederverwendung in verteilten Content Management Systemen ist wahrlich nicht leicht zu implementieren und stellt einige Herausforderungen dar, jedoch ist der Nutzen bei effektivem Einsatz unbezahlbar.

Literaturverzeichnis

[Alka03a] Alkacon Software, OpenCms Documentation Version 5.0, Version 1.0, 2003, http://www.opencms.org/export/download/opencms/opencms_5.0_doc_full.zip

[Alka03b] Alkacon Software, OpenCms 5.0 presentation slides, 2003, http://www.opencms.org/export/download/alkacon/Slides_OpenCms_5_Overview_Alkacon.pdf

[Baum01] S. Baumann, Der Content Management Kernprozess, 2001, http://www.contentmanager.de/magazin/artikel_61_der_content_management_kernprozess.html

[Berg03] W. Bergthaler, R. Buchgraber, R. Fankhauser, R. Galler, P. Germ, H. Gründhammer, Content Management, FH Joanneum, 2003

[Bitt04] T. Bittner, OpenCms, Hauptseminar Content Management Systeme, 2003/2004, http://wwwdb.informatik.uni-rostock.de/(en,nobg)/Lehre/Vorlesungen/hs-ws0304/OpenCMS.pdf

[Boik02] B. Boiko, Content Management Bible, 1. Auflage, Wiley Verlag, Indianapolis, 2002

[Ecom03] e-commerce online, Aktuelles: Content Management Systeme, 2003 http://www.e-commerce-online.de/Aktuelles/cms.htm

[Engl04] B. English, O. Londer, S. Shell, T. Bleeker, S. Cawood, Microsoft Content Management Server 2002: A Complete Guide, 2. Auflage, Addison Wesley Verlag, Boston, 2004

[Hack02] J. T. Hackos, Content Management for Dynamic Web Delivery, 1. Auflage, Wiley Verlag, New York, 2002

[Hage05] O. Hager, Content Management mit Typo3, 2005, `http://www.lizenzfrei.at/fileadmin/templates/vortrag.pdf`

[Inte05a] Interwoven, Interwoven TeamSite Content Server, Datasheet, 2005, `http://www.interwoven.com/documents/datasheets/dsteamsite65.pdf`

[Inte05b] Interwoven, Interwoven Content Transformation Services for TeamSite, Datasheet, 2005, `http://www.interwoven.com/documents/datasheets/dscts2.pdf`

[Klei02] C. Kleiner, T. Eppler, Klärung der Begrifflichkeiten und Abgrenzung von MAM gegenüber CM, DM und KM Systemen, 2002, `http://www.internetmanagement.ch/index.cfm/fuseaction/shownews/newsid/391/`

[Knap03] K. Knappe, Content Management Systeme – Ein Überblick zur Content Verwaltung sowie zum aktuellen Marktumfeld, TFH Berlin, 2003

[Manh03] K. Manhart, Die Website im Griff, Internet Magazin, 2003 `http://www.internet-magazin.de`

[Meye04] R. Meyer, Grundlagenwissen Typo3 Version 3.6.2: Deutsche Typo3-Dokumentation, Version 2.0, 2004, `http://www.mittwald.de/fileadmin/pdf/dokumentation.pdf`

[Müll05] T. Müller-Prothmann, H. Partl, Typo3 Einführung, 2005, http://www.boku.ac.at/typo3einf/typo3einf.pdf

[Naka02] R. Nakano, Web Content Management: A collaborative approach, 2. Auflage, Addison-Wesley Verlag, Indianapolis, 2002

[Pinu03] Pinuts media+science Multimedia-Agentur GmbH, Content Management Systeme Einführung, http://www.pinuts.de/WP_CMS_Einfuehrung.pdf

[Rapo04] J. Rapoza, TeamSite 6.1, 2004, http://www.eweek.com/article2/0,1759,1627987,00.asp

[Robe02] J. Robertson, Centralised or decentralised authoring, KM Column, 2002, http://www.steptwo.com.au/papers/kmc_decentralised/index.html

[Robe04] J. Robertson, Content reuse in practice, KM Column, 2004, http://www.steptwo.com.au/papers/kmc_contentreuse/index.html

[Rock03] A. Rockley, P. Kostur, S. Manning, Managing Enterprise Content: A Unified Content Strategy, 1. Auflage, Newriders Verlag, Indianapolis, 2003

[Schu00] E. Schuster, S. Wilhelm, Content Management, Informatik-Spektrum, Band 23, Nummer 6, Springer-Verlag, Berlin, Dezember 2000

[Schu04] K. Schulz, Typo3, Hauptseminar Content Management Systeme, 2004, http://wwwdb.informatik.uni-

rostock.de/Lehre/Vorlesungen/hs-ws0304/Typo3.pdf

[Schw04] A. C. Schwickert, Dezentrales Web Content Management, Arbeitspapiere WI, Nr. 5/2004, Justus-Liebig-Universität Gießen, 2004, http://www.web-portal-system.de/wps/transmit/dl/det/wsegmbh/144/

[Skar04] K. Skårhøj, R. Lemke, Extension Key: templavoila, 2004, http://typo3.org/fileadmin/pdf_manuals/manual-templavoila-09-09-2004_05-36-28.pdf

[Toms01] M. Tomsen, Killer Content – Strategien für das erfolgreiche Content Management im eCommerce, 1. Auflage, Addison-Wesley Verlag, München, 2001

[Trut04] Truthmedia, Using Interwoven, 2004, http://downloads.truthmedia.com/interwoven/usinginterwovencccc.ppt

[Wald04] D. Waldt, The trend towards distributed content management, Gilbane Report, Vol. 12, Nr. 2, 2004, http://gilbane.com/gilbane_report.pl/96/The_Trend_Towards_Distributed_Content_Management.html

[Weit01] L. Weitzman, S. Elo-Dean, D. Meliksetian, K. Gupta, N. Zhou, J. Wu, Transforming the content management process at IBM.com, Case studies of the CHI2002|AIGA Experience Design FORUM, Minneapolis, 2002

[Weit02] L. Weitzman, S. Elo-Dean, D. Meliksetian, J. Milton, N. Zhou, P. Davis, J. Wu, XML content management: Challenges and solutions, In Proceedings of XMLEurope, Berlin, 2001

Webseiten:

[Cmsm] CMS Vergleich, `http://www.cmsmatrix.org`

[Cont] Contentmanager.de CMS Vergleich, `http://www.contentmanager.de`

[Inte] Interwoven, `http://www.interwoven.com`

[Micr] Microsoft Content Management Server, `http://www.microsoft.com/cmserver`

[Open] OpenCms, `http://www.opencms.org`

[Typo] Typo3, `http://www.typo3.org`